IJIE (Indonesian Journal of Informatics Education)

Research Paper ISSN: 2549-0389

Implementation of K-Means Clustering for Optimization of Student Grouping Based on Index of Learning Styles in Programming Classes

*** Dwi Maryono**
Informatics Education Department
Universitas Sebelas Maret
dwimaryono@staff.uns.ac.id

Cucuk Wawan Budiyanto
Informatics Education Department
Universitas Sebelas Maret
cbudiyanto@staff.uns.ac.id

Allan Auri Pamungkas
Informatics Education Department
Universitas Sebelas Maret
allan.auri@gmail.com

Abstract:

This study aims to group students into study groups (classes) based on learning styles utilising K-Means Clustering technique and Sum of Squared Error for cluster assessment. This study used type of learning style developed by Felder and Silverman, which includes four dimensions: (1) the learning process; (2) perception of learning; (3) information input; and (4) understanding of information. This study subjects were Universitas Sebelas Maret's students majoring in informatics Education consisting of 58 respondents. The results showed that the K-Means clustering approach with cluster evaluation using Sum of Squared Error produced the best clustering when the number of clusters was k=2. The cluster analysis showed that each class has different learning styles and characteristics. The first cluster (group) consists of 26 respondents and has the features of an active learning style in the learning process dimension, sensing in the learning perception dimension, visual in the information input dimension, and a balance between global and sequential in the information understanding dimension. Meanwhile, the second cluster (group) consists of 32 respondents and has a reflective tendency in the learning process dimension, sensing in the learning perception dimension, visual in the information input dimension, and global in the information understanding dimension.

Keywords: Index of Learning Style, K-Means Clustering, Learning Styles Grouping.

DOI: http://dx.doi.org/10.20961/ijie.v6i2.68151

Introduction

Each individual has differentiations in the optimization of receiving information. The way to understand the characteristics of each individual in receiving information is by knowing their learning style. Learning styles are keywords for developing performance at work, school, and in interpersonal situations (Porter & Hernacki, 2015). Each individual has a different learning style according to the individual characteristics in the learning process. Each person's personality, which varies by their bearing, experience, education, and developmental history, is tightly related to their learning methods (Mulyono, 2012). In other words, because diverse circumstances significantly impact individual traits, each person has a unique learning style.

There are several ways to evaluate each person's learning preferences. One of the guidelines for knowing individual learning styles is the Index of Learning Styles (ILS) score (Soloman & Felder, 1999). The ILS classifies the learning style of each individual into four dimensions of the learning style (Felder & Silverman, 1988). Those dimensions are (1) learning process; (2) learning perception; (3) information input; and (4) understanding of information. From the learning perception dimension, students tend to be of the sensing type (concrete thinkers, practical, fact and procedure-oriented) or intuitive (abstract thinkers, innovative, theory-oriented and underlying meanings). From the information input dimension, students tend to be of the visual type (preferring visual representations presented by material, such as pictures, diagrams and flowcharts) or verbal (preferring written and spoken explanations); From the dimension of the learning process, students tend to be of the active type (learn by trying various things, enjoy working in groups) or reflective (learn by thinking things through, preferring to work alone or with a close partner); From the dimension of understanding information, students tend to be of the sequential type (linear thinking process, learning in a small number of additional steps) or global (holistic thinking process, learning in giant leaps).

The Bachelor of Informatics Education program is one of the study programs at Universitas Sebelas Maret, Indonesia which consists of two classes in every batch. The grouping of learning classes currently uses a classification system based on odd and even student identification numbers. This method of grouping student classes is ineffective because each student can process different information. Therefore, a categorization that considers the similarity of student learning styles will be more effective in conducting classes.

Grouping students based on learning styles can make it easier for teachers to design learning, such as determining the appropriate media, types of student activities, and selecting learning models and methods. In the field of programmers, several studies have shown that selecting the right learning strategy affects student learning outcomes. To anticipate differences in learning styles, the teacher should apply active approaches to motivate different kinds of learners (Maia et al., 2017). Pair programming, dialogues, and collaborative and cooperative learning are a few examples of learning methodologies (Othman et al., 2013; Tie & Umar, 2010; Umar & Hui, 2012). Other studies have shown a correlation between students and teacher learning styles (de Raadt & Simon, 2011). Students with the same learning style as the teacher tend to have better results than other students. As a result, it recommends that a teacher be able to meet the needs of pupils under their preferred methods of learning rather than only their own.

One technique to group a set of data according to the similarity of objects is to perform cluster analysis. Hair et al (Fadliana & Rozi, 2015) defines cluster analysis as a technique in multivariate statistical analysis that aims to group observation objects into groups based on their characteristics. By conducting a cluster analysis, we can determine student groups based on the similarity of student learning styles with each other. With this categorization, lecturers should be able to choose the best learning technique for each study group based on that group's learning style characteristics.

The cluster analysis technique used in this study was K-Means clustering. The first step in this K-Means clustering algorithm is to predict how many clusters will develop. In K-Means clustering, user usually chooses the number of clusters manually. However, users can also determine the optimal number of clusters based on specific criteria. We can use the sum of square error or silhouette calculation to find the optimal number of groups (Thinsungnoen et al., 2015). The optimal value of K will optimize the data so that the objects in any group are entirely homogeneous.

Based on the description above, we formulate some problems in this study, which are: 1) How many clusters are optimal in grouping Computer Education students based on ILS learning style; 2) How are the characteristics of the learning style of each group resulting from the cluster analysis; 3) What are the learning strategies based on clustering results in programming class?

Related Work

Previously, we researched student grouping based on the ILS learning style. We apply hierarchical clustering to group students into study groups (Pamungkas et al., 2021). However, in this previous study, we manually obtained the number of groups using graphs generated by the algorithm. There needs to be a measure to determine whether the number of groups produced is optimal. Therefore, in this study, we apply K-Means Clustering and SSE to obtain optimal grouping.

Research Method

This study used a descriptive quantitative method to describe the student's learning styles and study groups resulting from cluster analysis. The subject of this study was students of Bachelor of Informatics Education, Universitas Sebelas Maret year of 2019. Learning styles are measured using Felder and Solomon's Index of Learning Style (ILS) instrument. After students completed the questionnaire, the k-Means clustering algorithm served for k=2, 3, 4, and 8. From these values of k, we determined the value that optimised the clustering results using the Sum of Squared Error evaluation. A substantial knee, also known as an abrupt change in the SSE value, will signal the best k value. For example, in Figure 1, it can be seen that the significant value of k is k = 4 because there is a significant change from the previous values of k.

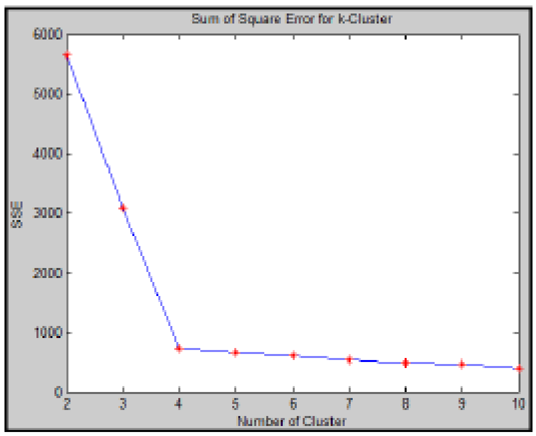

Figure 1. Example significant knee of SSE graphic

The clustering results are then analyzed based on the characteristics of the learning style of each cluster and used as a reference to determine the appropriate learning strategy for lecturers who will teach each cluster or class. We use Weka 3.6 software to implement K-Means Clustering

Result and Discussion

Result

a. Data Result

Respondents to the Index of Learning Styles questionnaire were 58 Bachelor of Informatics Education students, Universitas Sebelas Maret. Table 1 provides a summary of the data-gathering results.

Table 1. Descriptive Statistics of Overall ILS Learning Styles Data

	N	Minimum	Maximum	Mean	Std. Deviation
Active-Reflective	58	-9	11	-.55	4.418
Sensing-Intuitive	58	-5	11	3.31	3.724
Visual-Verbal	58	-7	11	4.97	4.312
Sequential-Global	58	-9	11	-1,21	3.741

b. K-Means Clustering Implementation

After the data is collected, the K-Means clustering algorithm applies with the values of k=2,3,4,5,6,7,8. Figure 2 displays the clustering outcomes for the value of k=2. The clustering results provide information on the SSE's magnitude, which is 8.65, the centroid of the two clusters, the average of the scores of each data dimension in each cluster, and the percentage of the number of members in each cluster. The properties of each cluster will be analysed subsequently using the centroid values since they represent the average score for each dimension.

```
Number of iterations: 2
Within cluster sum of squared errors: 8.650139408238367
Missing values globally replaced with mean/mode
Cluster centroids:
                        Cluster#
Attribute          Full Data        0         1
                     (58)       (26)       (32)
====================================================
Active-Reflective     -0.5517     3.3077    -3.6875
Sensing-Intuitive      3.3103     2.6154     3.875
Visual-Verbal          4.9655     6.0769     4.0625
Sequential-Global     -1.2069     0.7692    -2.8125

Time taken to build the Model (complete training data): 0.04 seconds
=== Model and evaluation on training set ===
Clustered Instances
0      26 ( 45%)
1      32 ( 55%)
```

Figure 2. K-Means clustering results for k=2.

The next step is comparing the magnitude of the SSE with the value of k to obtain the value of k that optimises clustering. Figure 3 displays the findings comparing the magnitude of SSE and k.

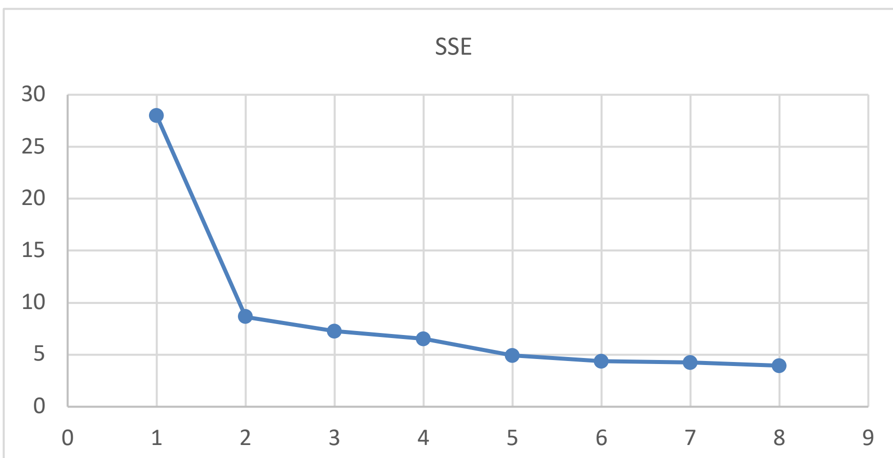

Figure 3. SSE graphics for k = 1, 2, 3,4, 5, 6, 7, 8.

Figure 3 shows that the significant knee is at k =2, so the number of the ideal groups in K-means clustering is 2. Therefore, from 59 students, it can be grouped into two classes as the result in Figure 2.

c. Clustering Result Analysis

From the previous results, we obtained that many optimal classes are k=2, so clustering can be used in Figure 2 to analyse each group. Figure 4 displays the average score (centroid) for the two groups' learning style components.

```
Attribute          Full Data        0         1

                     (58)       (26)       (32)

====================================================

Active-Reflective     -0.5517     3.3077    -3.6875

Sensing-Intuitive      3.3103     2.6154     3.875
```

Visual-Verbal	4.9655	6.0769	4.0625
Sequential-Global	-1.2069	0.7692	-2.8125

Figure 4. The centroids of the two clusters

The information processing dimension (active/reflective) becomes the most significant difference between clusters 0 and 1. The data on cluster 1 tends to have a reflective type with a score of -3.68. While cluster 0 tends to have an active type with a score of 3.30, distinguishing it from Cluster 1. In the learning perception dimension (Sensing-Intuitive), the data in clusters 0 and 1 have the same tendency, namely sensing on the learning perception dimension, with scores of 2.6154 and 3.875, respectively. Sensing types in learning perception tend to be more patient with details and experts in remembering facts and able to solve problems using proven methods (Wang et al., 2015). Sensibility-type learners favour learning material that has practical applications. In cluster 1, the tendency to learn to sense is higher than in cluster 0, but broadly speaking, both clusters have the same characteristics in learning perception. In the Information Input Dimension, both clusters show dominance, which has a visual tendency proving they are easier to capture visual information (images) than verbal (words/writings). So that learning that displays graphics, pictures, and diagrams becomes a selection of the right learning strategy so that student learning outcomes can improve. The data showed the visual learning style of cluster 0, with a score of 6.0769, more potent than cluster 1, which scored 4.0625. In the Information Understanding Dimension (Sequential-Global), cluster 0 has a score close to 0, which is 0.7692, so its position in the middle between sequential and global shows a balance in the dimension of understanding information. Meanwhile, cluster 1 shows respondents have a global trend with a score of -2.8125.

Discussion

The similarity of learning styles between the two clusters is in the second dimension, that is, students tend to be of the sensing type, and in the third aspect, where students tend to be of the visual type. Students with sensing type tend to be more patient with details, are experts in remembering facts, and can solve problems using proven methods. Students are passionate about learning things that have practical applications. Related to student programming, materials with this type will be suitable to apply problem- or project-based learning with a contextual approach, raising everyday problems. On the other hand, both clusters show a tendency of visual type in the third dimension. Therefore, the learner must present the material visually, not only text-based. To deliver the material, the teacher must add drawings, diagrams, or flowcharts. Students' comprehension of the programming content will benefit greatly from this.

The first dimension is where there is the most distinction between the two, where cluster 0 is more active while cluster 1 is more reflective. Active students learn by trying and experimenting, while reflective students learn by thinking about something (Felder & Silverman, 1988). Active students are much more comfortable and good at saying things on the spot and asking questions immediately, which aligns with the discussion activities. Therefore, learning should involve more physical activity and discussion. Students are very interested in the learning process in which they actively participate. Students actively learn best when they talk about the information obtained, experiment with it, or test it.

Reflective learners, on the other hand, tend to prefer to learn individually rather than together. They want to think about and rethink the information as soon as they learn something new. Reflective students inquire about their assumptions and engage in critical thought. The results showed that the K-Means clustering approach and cluster evaluation using The Sum of Squared Error produced the best clustering when the number of clusters was k=2. (SSE).

Therefore, in carrying out learning, teachers must give more time for reflective students to think about concepts, ideas, and problems in learning. They will carefully think about problem-solving strategies before coding practices in programming learning. Therefore, the provision of problems can be a little in quantity, a small amount. However, it covers a more extensive scale, allowing pupils adequate time to reflect, strategise, and implement. If discussion is needed, then there is no need to involve many students. One of the group learning approaches ideal for students with these traits is the think-pair-share Model. Two students are all needed for this, providing each student time to think before reflecting with his companion.

Conclusion

From the discussion results, BIE students may be in separate learning groups (classes) using the K-Means clustering technique. It can be inferred, with many optimal classes being k = 2. The results of cluster analysis showed that each class has different learning styles. Cluster 0, consisting of 26 respondents, tends to aspects of active learning processes, perceptions of learning sensing, visual inputting of information, and balance in aspects of understanding information, between global and sequential. Cluster 1, consisting of 32 respondents, tends to aspects of reflective learning processes, perceptions of sensing learning, visual inputting of information, and understanding learning globally. From the characteristics of each cluster, appropriate learning strategies can be designed, especially in programming lessons.

References

de Raadt, M., & Simon. (2011). My Students Do not Learn the Way I Do. *Conferences in Research and Practice in Information Technology Series*, *114*(Ace), 105–112.

Fadliana, A., & Rozi, F. (2015). Penerapan Metode Agglomerative Hierarchical Clustering Untuk Klasifikasi Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana. *CAUCHY. Jurnal Matematika Murni Dan Aplikasi*, *4*(1), 35–40.

Felder Richard M., N. C. S. U., & Silverman Linda K., I. for the S. of A. D. (1988). Learning and Teaching Styles in Engineering Education. *Engr. Education*, *78(7)*(January 1988), 674–681.

Maia, M. C. O., Serey, D., & Figueiredo, J. (2017). Learning styles in programming education: A systematic mapping study. *Proceedings - Frontiers in Education Conference, FIE*, *2017-October*(November 2019), 1–7. https://doi.org/10.1109/FIE.2017.8190465

Mulyono. (2012). *Strategi Pembelajaran*. UIN Maliki Press.

Othman, M., Othman, M., & Hussain, F. M. (2013). Designing Prototype Model of an Online Collaborative Learning System for Introductory Computer Programming Course. *Procedia - Social and Behavioral Sciences*, *90*(InCULT, 2012), 293–302. https://doi.org/10.1016/j.sbspro.2013.07.094

Pamungkas, A. A. P., Maryono, D., & Budiyanto, C. W. (2021). Cluster Analysis for Student Grouping Based on Index of Learning Styles. *IOP Conference Series: Earth and Environmental Science*, *1808*(1). https://doi.org/10.1088/1742-6596/1808/1/012023

Porter, B. D., & Hernacki, M. (2015). *Quantum Learning: Membiasakan Belajar Nyaman dan Menyenangkan (Terjemah)*. Kaifa Learning.

Soloman, B. A., & Felder, R. M. (1999). *Index of Learning Styles Questionnaire*. North Carolina State University. http://www.engr.ncsu.edu/learningstyles/ilsweb.html

Thinsungnoen, T., Kaoungku, N., Durongdumronchai, P., Kerdprasop, K., & Kerdprasop, N. (2015). *The Clustering Validity with Silhouette and Sum of Squared Errors*. 44–51. https://doi.org/10.12792/iciae2015.012

Tie, H. H., & Umar, I. N. (2010). The impact of learning styles and instructional methods on students' recall and retention in programming education. *Proceedings of the 18th International Conference on Computers in Education: Enhancing and Sustaining New Knowledge Through the Use of Digital Technology in Education, ICCE 2010*, *January*, 191–195.

Umar, I. N., & Hui, T. H. (2012). Learning Style, Metaphor and Pair Programming: Do they Influence Performance? *Procedia - Social and Behavioral Sciences*, *46*(December 2012), 5603–5609. https://doi.org/10.1016/j.sbspro.2012.06.482

Wang, Y., Han, X., & Yang, J. (2015). Revisiting the blended learning literature: Using a complex adaptive systems framework. *Educational Technology and Society*, *18*(2), 380–393.