e-ISSN:
JURNAL SISTEM DAN TEKNOLOGI INFORMASI
The Journal of Information System and Technology
https://journal.uhamka.ac.id/index.php/sistekinfo/

UHAMKA PRESS

Volume 2 • Number 1 • June 2025 • 1 - 9

Student Learning Style Segmentation Using K-Means Clustering for Adaptive Learning in Higher Education

Izhar Syafawi[1], Selphia Nur Azzahra[2], Mohammad Givi Efgivia [3]
[1]Informatics Engineering, Muhammadiyah of Prof. Dr. HAMKA University, Indonesia
[2]Informatics Engineering, Muhammadiyah of Prof. Dr. HAMKA University, Indonesia
[3]Information Technology System, Prof. Dr. Hamka Muhammadiyah University, Jakarta, Indonesia

Received: June 8, 2025 Accepted: June 8, 2025 Published: June 8, 2025

Abstract

Understanding students' learning styles is crucial in developing adaptive instructional strategies that enhance engagement and academic achievement in higher education. This study aims to classify university students based on their dominant learning styles—visual, auditory, and kinesthetic—using the K-Means clustering algorithm. A total of 150 respondents participated in a questionnaire designed with Likert-scale items measuring learning preferences. The clustering process successfully segmented the students into three well-defined groups: visual learners (38.7%), auditory learners (31.3%), and kinesthetic learners (30%). Cluster validation using the Silhouette Score yielded an average value of 0.62, indicating a strong internal consistency and effective group separation. Further analysis revealed that 75% of students in the visual cluster showed higher comprehension using visual materials such as diagrams and videos. In the auditory cluster, 70% preferred oral explanations, discussions, and lectures, while 65% of kinesthetic learners performed better through hands-on activities and physical interaction with learning materials. These insights emphasize the importance of integrating personalized approaches into course design to support diverse learning needs. The findings suggest that adaptive learning models based on clustering techniques like K-Means can assist educators in tailoring instruction, thereby improving the overall learning experience. This approach supports more inclusive and effective pedagogy in higher education environments..

Keywords: made up 3-5 keywords

INTRODUCTION

According to the United Nations Educational, Scientific and Cultural Organization (UNESCO), each educational institution must be able to develop four pillars of education for both the present and the future, namely: learning to know, learning to do—where students are required to be skilled in doing something—learning to be, and learning to live together. Learning serves as a means for individuals to acquire and develop new knowledge, skills, abilities, behaviors, and attitudes; to become skilled from being unskilled, and to be able to do something from not knowing how to do it. All of this is achxieved through experience or intentional interaction with the environment.(Lailiyah et al., 2019)

In modern educational systems, especially at the higher education level, there is a growing recognition of the need to move beyond a one-size-fits-all teaching model. Students come from diverse backgrounds, possess different cognitive strengths, and respond uniquely to various instructional methods. This diversity makes it essential to consider how students learn most effectively. An effective learning process is strongly influenced by the individual learning styles of students. A learning style refers to a person's characteristic way of absorbing, processing, and understanding information. Each student may prefer a different mode of learning, such as visual, auditory, or kinesthetic.

However, in higher education practices, teaching methods tend to be uniform and often fail to consider the diversity of students' learning styles. This may result in some students struggling to understand the material, as it does not align with how they learn best. When students' preferred learning styles are not acknowledged, their engagement, comprehension, and academic performance may be negatively impacted.

Recognizing this gap, it becomes crucial for educators and academic institutions to adopt more student-centered approaches. One of the strategies to achieve this is by classifying or segmenting students based on their dominant learning styles. Such segmentation enables educators to tailor their instructional designs, teaching media, and interaction methods according to the needs of different student groups.

To facilitate this segmentation process efficiently and accurately, data-driven methods such as K-Means Clustering can be utilized. K-Means is a widely used unsupervised learning algorithm in data mining that can group individuals based on the similarity of their attributes—in this case, learning preferences. Through this approach, a deeper understanding of the student population can be obtained, leading to the development of more personalized learning environments that are aligned with students' actual needs.

Ultimately, integrating data mining techniques like K-Means into educational analysis not only supports pedagogical improvement but also contributes to the broader goal of inclusive, adaptive, and effective higher education practices. This research, therefore, aims to explore student segmentation based on learning styles using K-Means clustering and to highlight its implications for instructional strategy design in university settings.

LITERATUR RIVIEW

K-Means Clustering

In K-means clustering is one of the most well-known and widely used unsupervised learning methods in pattern recognition and machine learning. Although categorized as an unsupervised method, the algorithm requires an initial specification of the number of clusters, making it not entirely unsupervised. In response to this limitation, recent research has introduced U-KMeans or Unsupervised K-Means, which no longer depends on initial parameter settings or cluster initialization. This algorithm can automatically determine the optimal number of clusters without manual intervention. Experimental results and comparisons have demonstrated that the U-KMeans algorithm delivers better performance in data clustering compared to conventional methods.(Bhatia, n.d.)

Furthermore, Lloyd's algorithm, a popular implementation of K-Means clustering, has been widely applied in various fields such as color quantization, data compression, and segmentation. This algorithm operates by finding cluster centers that minimize the average squared distance from each data point to its nearest center. One efficient variation of this method is the filtering algorithm, which utilizes a kd-tree data structure to accelerate the search process. Experimental studies show that the filtering algorithm is highly efficient in processing large datasets, whether synthetic or real, making it suitable for applications that require fast computation and accurate clustering results.(Kanungo et al., 2000)

On the other hand, other research highlights one of the major challenges in K-Means clustering: its dependence on the initialization of cluster centers, which can significantly affect the quality of the clustering results. To address this issue, an adaptive approach has been developed that enables clusters to grow dynamically without relying on initial cluster representations. This technique allows for the creation and merging of clusters flexibly throughout the process while maintaining a constant number of clusters. This approach has proven effective in accelerating data search processes in situations where other efficient search techniques are not available, and in improving segmentation accuracy.(Sinaga & Yang, 2020)

Student Segmentation

Student-teacher frameworks have emerged as a powerful strategy in tackling visual anomaly detection, especially in one-class classification and segmentation tasks. Recent research introduced DeSTSeg, an enhanced student-teacher-based segmentation model that integrates a pre-trained teacher network with a denoising student encoder-decoder and a segmentation module. By training the student to replicate teacher features from synthetically corrupted normal images, the model enforces robustness in representation learning. Furthermore, the addition of synthetic anomaly masks for segmentation supervision allows DeSTSeg to adaptively fuse multi-level features, yielding state-of-the-art results on industrial inspection benchmarks.(Huo et al., n.d.)

In the domain of semi-supervised semantic segmentation, the challenge of intra-class variation—where similar classes exhibit varied visual patterns—has been addressed through feature regularization in student-teacher structures. One notable approach combines a linear predictor and a prototype-based predictor to ensure intra-class consistency, effectively tightening the distribution of features from the same pseudo-class. This architecture, when enhanced with CutMix data augmentation and a dynamic prototype updating strategy, facilitates robust label propagation in pixel-level predictions. The method has shown superior performance on widely-used datasets such as Pascal VOC and Cityscapes.(Zhang et al., n.d.)

Another advancement in student-teacher segmentation is the Asynchronous Teacher-Student Optimization (ATSO) algorithm, which targets the issue of *lazy mimicking*—a tendency of models to retain outdated predictions. ATSO overcomes this limitation by alternating the learning and labeling phases across subsets of unlabeled data. This asynchronous update mechanism refreshes the pseudo-labels more effectively, enhancing the generalization capability of the student network. Evaluations across both medical and natural image segmentation benchmarks affirm the method's competitiveness against existing state-of-the-art models, especially in low-annotation and domain adaptation scenarios.(Xu et al., n.d.)

Learning Style

Learning styles play a crucial role in enhancing student engagement and personalizing instruction within adaptive e-learning environments. Recent research has emphasized the design of adaptive e-learning systems that align with individual learning preferences to optimize educational outcomes. By employing mixed methods—including development-based design and quasi-experimental research—one study demonstrated that students exposed to a learning-style-based adaptive environment showed significantly higher engagement across affective and behavioral factors compared to those in a conventional e-learning setting. These findings highlight the importance of tailoring digital learning environments to student learning styles to boost emotional and participatory involvement.(Hassan et al., 2021)

Despite the flexibility and accessibility that e-learning platforms offer, they often suffer from high drop-out rates, primarily due to the lack of personalized learning experiences. While gamification has been explored as a strategy to increase engagement, it falls short in generating sustained intrinsic motivation among diverse learners. To address this, a novel framework has been introduced that adapts gamification elements to users' identified learning styles—determined through system interaction. Experimental results reveal that this adaptive gamification approach can significantly improve learner motivation (by 25%) and reduce drop-out rates (by 26%), underscoring the value of integrating learning styles into system design for deeper user engagement.(El-Sabagh, 2021)

Moreover, understanding learning styles has become a foundational element in educational research, with numerous pedagogic studies attempting to classify and apply different styles in academic contexts. Scholars have argued that acknowledging the diversity in how students absorb and process information can enrich learning experiences and outcomes. However, the literature also cautions about the limitations of rigidly applying learning styles without accommodating evolving educational contexts. The discussion around the feasibility, effectiveness, and adaptability of learning styles continues to influence the development of more nuanced and flexible instructional strategies in both traditional and digital education.(Yadav & Shukla, 2021)

METHODS

This study employs a quantitative research approach, integrating data mining techniques to achieve objective and replicable results. Specifically, the K-Means clustering algorithm is used to segment students based on their individual learning styles. The overall goal of this methodological framework is to identify natural groupings within the student population that reflect distinct preferences in how they process and absorb information. The use of a clustering algorithm in this context provides a systematic and data-driven way to uncover patterns that might not be immediately visible through traditional analysis.

The research process begins with the design and administration of a structured questionnaire. This questionnaire is carefully developed to assess students' learning preferences across three major modalities: visual, auditory, and kinesthetic. The items are constructed based on well-established learning style theories and are presented using a Likert scale, typically ranging from "strongly disagree" to "strongly agree." This scale allows for the quantification of each student's inclination toward specific learning behaviors and activities. For example, a visual learner might strongly agree with statements like "I remember things better when I see pictures or diagrams," while a kinesthetic learner may resonate more with statements such as "I understand better when I engage in hands-on activities."

Once a sufficient number of responses are collected, the data enters the preprocessing stage. This stage is essential to ensure data quality and relevance before conducting clustering analysis. First, data cleaning is performed to eliminate incomplete, inconsistent, or invalid responses, such as missing values or contradictory answers. This process ensures that only reliable and accurate data are used in further analysis, thus enhancing the credibility of the findings.

Following cleaning, the dataset undergoes normalization, a process that standardizes all variables to a common scale. Since different items in the questionnaire may have varying ranges, normalization is crucial to prevent disproportionate influence from any single variable. This step not only ensures a fair comparison between data points but also significantly improves the effectiveness of the K-Means algorithm, which is sensitive to the scale of input features.

Next, it is necessary to determine the optimal number of clusters (K) that should be used in the K-Means algorithm. To achieve this, the study applies the Elbow Method, a well-known technique in cluster analysis. This method involves running the K-Means algorithm multiple times with different values of K (e.g., from 1 to 10) and computing the Within-Cluster Sum of Squares (WCSS) for each iteration. The WCSS indicates the compactness of the clusters—the lower the WCSS, the more cohesive the clusters. The point on the graph where the reduction in WCSS starts to plateau (forming an "elbow" shape) is identified as the optimal number of clusters. This step ensures that the segmentation is neither too generalized nor too fragmented.

After determining the most suitable value for K, the K-Means algorithm is executed on the normalized dataset. The algorithm starts by initializing K centroids randomly. Each student (data point) is then assigned to the nearest centroid based on Euclidean distance. The centroids are recalculated in each iteration as the mean of all data points in the cluster. This iterative process continues until the assignments no longer change significantly, or until a predefined number of iterations is reached. The final output consists of distinct and well-formed clusters representing groups of students with similar learning preferences.

Once clustering is complete, each group is analyzed to interpret the characteristics that define it. This involves examining the average responses within each cluster to identify dominant learning styles. For instance, one cluster may show a higher average score on visual-related questions, while another may lean toward auditory or kinesthetic preferences. This interpretive phase is critical for translating statistical groupings into meaningful educational insights.

To validate the quality of the clustering results, internal validation techniques are applied—most notably the Silhouette Score. This metric assesses how similar a data point is to its own cluster compared to other clusters. A higher average silhouette score indicates better-defined clusters, reflecting a clearer separation between different learning style groups and greater internal coherence within each cluster. This validation step adds rigor to the analysis and increases confidence in the reliability of the segmentation.

In summary, this methodological framework combines well-established data preprocessing techniques, systematic clustering procedures, and rigorous validation metrics to ensure the robustness and accuracy of the study. Through this approach, the research aims to generate actionable insights into students' learning preferences and provide a solid foundation for educators to implement differentiated instruction strategies that cater to the unique needs of each student group.

FINDINGS AND DISCUSSION

After applying the K-Means clustering algorithm to the learning style data collected from 150 students, three primary clusters were identified, each representing groups of students with distinct learning style characteristics. The distribution of students across these clusters was relatively balanced, with Cluster 1 comprising 58 students (38.7%), Cluster 2 containing 47 students (31.3%), and Cluster 3 including 45 students (30%). Each cluster revealed a dominant learning style preference, namely visual, auditory, and kinesthetic.

The first cluster was dominated by students with a visual learning style. These students tend to understand and retain information more effectively when it is presented through visual media, such as images, diagrams, infographics, and educational videos. Approximately 75% of students in this cluster reported that they find it easier to comprehend material when it is delivered visually rather than through text or spoken explanations. This highlights the importance of utilizing visual-based instructional methods to enhance the learning experience for this group. Teaching strategies that incorporate well-designed presentation slides, mind maps, concept illustrations, and multimedia content are recommended to meet the needs of visual learners.

The second cluster consisted of students who demonstrated a preference for auditory learning. About 70% of the students in this group stated that they grasp information more effectively through verbal explanations, discussions, and audio content. These learners tend to benefit most from hearing information, whether through live lectures, peer discussions, or recorded materials such as podcasts. To optimize learning outcomes for this cluster, instructional approaches should include interactive lectures, question-and-answer sessions, group discussions, and opportunities for oral presentations. Such methods not only support comprehension but also encourage active participation and collaboration among students.

The third cluster included students with a kinesthetic learning style. Around 65% of this group indicated a preference for hands-on learning through physical activities, experiments, or simulations. These students tend to learn best when they are actively engaged in the learning process through direct experience. Therefore, teaching strategies that involve project-based learning, lab activities, real-life problem-solving tasks, and physical demonstrations are strongly recommended for this cluster. Incorporating movement and tactile experiences into lessons can significantly improve knowledge retention and engagement among kinesthetic learners.

To validate the quality of the clustering results, a Silhouette Score was calculated, yielding an average score of 0.62. This value suggests a reasonably good separation between clusters and high intra-cluster similarity among members. In other words, the clustering can be considered valid and reliable as a foundation for developing targeted and differentiated teaching strategies that align with students' learning preferences.

In conclusion, this study reinforces the understanding that students possess diverse learning needs, and a one-size-fits-all approach to instruction is unlikely to be effective. Educators are encouraged to design flexible and adaptive learning experiences that cater to various learning styles. By utilizing the insights from clustering analysis, personalized learning strategies can be developed to improve learning outcomes, increase student engagement, and foster a more inclusive educational environment. This approach has the potential to significantly enhance the overall quality of education, especially in today's digitally-driven and learner-centered landscape.

CONCLUSION

Based on the analysis of learning style data using the K-Means clustering algorithm on a sample of 150 students, this study successfully categorized students into three main clusters, each representing a distinct learning style: visual, auditory, and kinesthetic. This segmentation reveals that every student tends to have a unique way of processing and understanding information, which significantly affects their learning outcomes. With 38.7% of students in the first cluster, 31.3% in the second, and 30% in the third, the distribution demonstrates a relatively balanced presence of each learning style across the population studied.

The first cluster is composed mainly of students with a visual learning style, who prefer learning through visual aids such as images, diagrams, concept maps, and educational videos. These students reported a higher ability to remember and comprehend information when it is presented visually. This suggests that integrating visual elements into instructional materials, such as infographics, animations, and visually engaging presentations, is crucial to support their learning process effectively.

The second cluster includes students with an auditory learning style. These learners tend to understand material better through listening and verbal interactions, such as lectures, discussions, or audio recordings. The findings indicate that incorporating verbal-based teaching strategies—such as interactive lectures, peer discussions, and educational podcasts—will enhance learning effectiveness for this group.

The third cluster comprises students with a kinesthetic learning style, who prefer learning through physical activities and hands-on experience. They show improved comprehension when actively involved in simulations, experiments, or real-world problem-solving tasks. Therefore, applying experiential learning methods such as project-based learning, laboratory activities, and fieldwork is highly recommended for this group.

To validate the clustering results, the Silhouette Score was used and yielded an average score of 0.62. This value indicates a fairly good separation between clusters, with high internal similarity among cluster members. Thus, the clustering outcome is considered both valid and reliable, making it a strong foundation for designing personalized learning strategies.

In conclusion, this study highlights the diverse learning preferences among students and emphasizes the need for educators to adopt a differentiated teaching approach. A uniform method of instruction cannot effectively meet the needs of all learners. Therefore, educational institutions and instructors are encouraged to develop adaptive learning environments that consider these clustered learning styles. By combining visual media, auditory engagement, and kinesthetic activities, educators can create more inclusive and effective learning experiences. Ultimately, this tailored approach can enhance student motivation, participation, and academic achievement across various learning styles.

REFERENCES

Bhatia, S. K. (n.d.). Adaptive K-Means Clustering. www.aaai.org

El-Sabagh, H. A. (2021). Adaptive e-learning environment based on learning styles and its impact on development students' engagement. International Journal of Educational Technology in Higher Education, 18(1). https://doi.org/10.1186/s41239-021-00289-4

Hassan, M. A., Habiba, U., Majeed, F., & Shoaib, M. (2021). Adaptive gamification in e-learning based on students' learning styles. Interactive Learning Environments, 29(4), 545–565. https://doi.org/10.1080/10494820.2019.1588745

Huo, X., Xie, L., He, J., Yang, Z., Zhou, W., Li, H., & Tian, Q. (n.d.). ATSO: Asynchronous Teacher-Student Optimization for Semi-Supervised Image Segmentation.

Kanungo, T., Mounv, D. M., Silverman~, R., Netanyahu, N. S., Wu, A. Y., & Piatko~, C. (2000). The Analysis of a Simple k-Means Clustering Algorithm. In Hong Kong China Copyright ACM.

Lailiyah, S., Yulsilviana, E., & Andrea, R. (2019). Clustering analysis of learning style on anggana high school student. Telkomnika (Telecommunication Computing Electronics and Control), 17(3), 1409–1416. https://doi.org/10.12928/TELKOMNIKA.V17I3.9101

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. IEEE Access, 8, 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

Xu, H.-M., Liu, L., Bian, Q., & Yang, Z. (n.d.). Semi-supervised Semantic Segmentation with Prototype-based Consistency Regularization. https://github.com/HeimingX/semi_seg_proto.

Yadav, S., & Shukla, G. (2021). Learning styles: A detailed literature review. International Journal of Applied Research, 7(2), 297–305. https://doi.org/10.22271/allresearch.2021.v7.i2e.8291

Zhang, X., Li, S., Li, X., Huang, P., Shan, J., & Chen, T. (n.d.). DeSTSeg: Segmentation Guided Denoising Student-Teacher for Anomaly Detection