2_Similarity 113 Fauka

by Turnitin No Repository

Identifying Student Learning Behavior Patterns Using K-Means Clustering (A Case Study At STMIK Mardira Indonesia)

Fitri Rizqiati[1],Ahfi Fauka Pranata[2]
STMIK Mardira Indonesia, Bandung[1,2]
fitririzqiati@gmail.com[1], ahfifauka.id@gmail.com[2]

Understanding student learning behavior is essential for developing effective instructional strategies and improving academic evaluation systems in higher education. This study aims to identify and characterize student learning behavior patterns using a clustering approach based on academic assessment data recorded in the Academic Information System (SIAKAD). To capture stable and long-term learning behavior tendencies, this research utilizes longitudinal academic records collected over eight consecutive semesters. The analyzed learning behavior attributes include assignment scores, quiz results, midterm examination scores, final examination scores, and attendance rates. The K-Means clustering algorithm was applied following data preprocessing and z-score standardization, while the optimal number of clusters was determined using the silhouette coefficient. The results reveal three distinct learning behavior patterns, namely students with low learning engagement, students with moderate engagement characterized by an exam-oriented learning strategy, and students with high and consistent learning engagement across learning activities.
Keyword: *Learning Analytics, Student Behavior, Clustering, K-Means, Academic Data*

INTRODUCTION

The widespread adoption of Academic Information Systems in higher education institutions has led to the accumulation of extensive and diverse academic datasets. These datasets encompass various academic components, including assignment scores, quiz results, midterm examinations, final examinations, and student attendance records collected over multiple semesters. Previous studies have indicated that such institutional academic data offer significant opportunities for learning analytics and educational data mining beyond their conventional administrative use (Utami et al., 2024; Zainuddin & Almuhammadi, 2023). Nevertheless, in many higher education institutions, academic data are still predominantly utilized for administrative purposes and summative academic decision-making, where final grades are treated as the primary indicator of student achievement.

This outcome-focused evaluation practice gives rise to an important academic phenomenon in which students exhibiting different levels of learning engagement and learning strategies may attain comparable academic outcomes. Several recent studies have reported that summative assessment systems tend to obscure variations in student learning behavior and engagement throughout the learning process (Ouassif et al., 2025; Franke et al., 2024). As a consequence, learning processes experienced by students over extended periods are not adequately represented in traditional academic evaluation frameworks.

From a theoretical standpoint, contemporary learning theories underscore the role of sustained engagement, learning consistency, and formative assessment in fostering meaningful and long-term learning experiences. However, a persistent theoretical gap remains between process-oriented learning principles and outcome-oriented academic evaluation practices. This gap has been widely discussed in the learning analytics literature, which emphasizes that learning processes should be examined alongside learning outcomes to achieve a more comprehensive and longitudinal understanding of student learning behavior (Zainuddin & Almuhammadi, 2023; Chen, 2025).

Empirical research in educational data mining and learning analytics, particularly within the Indonesian higher education context, demonstrates a growing interest in utilizing academic data to analyze student learning behavior. Utami et al. (2024) employed unsupervised learning methods to cluster student learning behaviors based on learning activity data, revealing that behavioral patterns can be systematically identified using institutional data sources. Likewise, Putri et al. (2023) showed that K-Means clustering is effective in grouping educational indicators and uncovering latent structures within educational datasets.

In contrast, many existing studies have concentrated on supervised learning approaches to predict academic performance and student dropout risk (Santosa et al., 2021; Hidayat, 2022). Although such approaches are beneficial for early academic intervention, recent reviews suggest that predictive models offer limited insight into the diversity and long-term stability of student learning behavior patterns (Franke et al., 2024; Ouassif et al., 2025). Consequently, variations in learning strategies among academically active students may not be adequately captured.

As a result, clustering-based techniques have gained increasing attention as exploratory approaches for identifying latent learning behavior patterns using multi-dimensional academic data. Previous studies have shown that K-Means clustering can effectively reveal heterogeneous student profiles by integrating academic assessment components and attendance data (Utami et al., 2024; Zainuddin & Almuhammadi, 2023). However, many of these studies remain constrained to short-term analyses, course-level data, or single indicators such as GPA, thereby limiting their ability to represent holistic and long-term learning behavior patterns.

Based on these research gaps and previous findings, the present study focuses on analyzing student learning behavior patterns using longitudinal academic data recorded in the Academic Information System. This study aims to identify learning behavior patterns derived from aggregated academic assessment components and attendance data collected over multiple semesters, as well as to examine the characteristics of each identified pattern. Furthermore, this research seeks to explore how different learning behavior patterns reflect variations in learning engagement over time, thereby providing insights to support more comprehensive academic evaluation and instructional decision-making in higher education.

METHOD

This study employed a quantitative research design with an exploratory approach to identify student learning behavior patterns based on academic data without predefined labels or categories. The exploratory approach was selected because the research aimed to uncover latent structures within academic data rather than to test causal relationships or predict academic outcomes. This study falls within the scope of Learning Analytics and Educational Data Mining, focusing on the analysis of learning processes as reflected in academic assessment components accumulated over multiple semesters.

An unsupervised learning technique, specifically K-Means clustering, was applied to group students based on similarities in their learning behavior characteristics. The research was non-experimental in nature, as no treatment or instructional intervention was applied to the research subjects. The clustering approach was chosen to explore naturally occurring learning behavior patterns within longitudinal academic data.

The population of this study consisted of students at the higher education institution under study whose academic assessment records were available across multiple semesters. To ensure that the identified learning behavior patterns reflected stable and long-term characteristics, the analysis utilized academic assessment data collected over the most recent eight consecutive semesters. This longitudinal approach was adopted to minimize short-term academic fluctuations and capture consistent learning behavior tendencies.

A total sampling technique was applied, in which all students who met the research criteria were included as the research sample. The inclusion criteria consisted of the availability of complete academic records across the selected semesters and the presence of all required assessment components in the Academic Information System. Records containing missing values in the selected variables were excluded during the preprocessing stage to ensure data completeness.

Data collection in this study utilized secondary data obtained from the Academic Information System. The collected data represented students' learning activities across multiple semesters and included aggregated assignment scores, quiz scores, midterm examination scores, final examination scores, and attendance rates. All data were anonymized previous to analysis to ensure confidentiality and ethical compliance.

Data analysis was conducted through several sequential stages. First, data preprocessing was performed, including data cleaning, completeness checking, and conversion of variables into numerical format. Second, all variables were standardized using z-score normalization to eliminate scale differences and ensure equal contribution of each variable in the clustering process. Third, the optimal number of clusters was determined using the silhouette coefficient, considering both statistical validity and pedagogical interpretability. Based on this evaluation, a three-cluster solution was selected as the most appropriate representation of the underlying learning behavior patterns. Fourth, K-Means clustering was applied to identify student groups with similar learning behavior characteristics. Fifth, cluster profiling was conducted by calculating the mean values of each variable to characterize the learning behavior attributes of each cluster.

RESULTS AND DISCUSSION

This section presents and discusses the results of the clustering analysis conducted to identify student learning behavior patterns using longitudinal academic data. The results are presented through tables and visualizations derived from statistical analysis, followed by a discussion that addresses the research objectives, explains how the findings were obtained, interprets the identified learning behavior patterns, and relates the results to established knowledge in learning analytics

The primary objective of this study was to identify student learning behavior patterns based on academic assessment components and attendance data collected over multiple semesters. The clustering results indicate that students can be grouped into three distinct learning behavior patterns, reflecting different levels and strategies of learning engagement over time. The optimal number of clusters was determined using the silhouette coefficient, with the three-cluster solution yielding the highest silhouette value and demonstrating adequate cluster separation for longitudinal behavioral data.

The distribution of students across the three clusters is presented in Table 1. The results show that students are unevenly distributed across clusters, suggesting natural variation in long-term learning behavior patterns rather than artificially balanced groupings.

Table 1. Distribution of Students by Cluster

Cluster	Pattern	Number Of Students
Cluster 0	Low Learning Engagement	1280
Cluster 1	Exam-Oriented Engagement	1105
Cluster 2	High and Consistent Engagement	347

This distribution indicates that while most students demonstrate moderate to high engagement levels, a smaller proportion consistently exhibits low engagement across academic components

To further explain how the clusters were formed, cluster profiling was conducted by analyzing the mean values of each learning behavior attribute for each cluster. Instead of presenting detailed numerical values, the profiling results are summarized in Table 2 in the form of dominant learning behavior characteristics. This approach was adopted to improve interpretability and readability, while detailed quantitative differences among clusters are illustrated through graphical visualizations.

Table 2. Mean Values of Learning Behavior Attributes by Cluster

Cluster	Learning Behavior Profile
Cluster 0	Low engagement across assignments, assessments, and attendance
Cluster 1	Moderate engagement with stronger emphasis on examinations
Cluster 2	High and consistent engagement across all learning components

Table 2 provides a concise overview of the learning behavior profiles identified through the clustering analysis. Cluster 0 represents students with consistently low engagement across assignments, assessments, and attendance, indicating limited participation throughout the learning process. In contrast, Cluster 2 demonstrates high and consistent engagement across all learning components, reflecting disciplined and sustained learning behavior over multiple semesters. Cluster 1 occupies an intermediate position, characterized by moderate overall engagement with a stronger emphasis on examination performance compared to continuous assessment components, suggesting an exam-oriented learning strategy.

These summarized profiles indicate that differences in student learning behavior are not solely reflected in overall performance levels but also in how learning engagement is distributed across academic activities. To further examine these differences, visual analysis was conducted using radar charts and centroid bar charts, which provide a detailed comparison of learning behavior attributes across clusters.

To support the interpretation of cluster characteristics, learning behavior profiles were visualized using a radar chart, as shown in Figure 1. The radar chart illustrates the normalized mean values of each learning behavior attribute for the three clusters, enabling a holistic comparison of engagement patterns across learning components

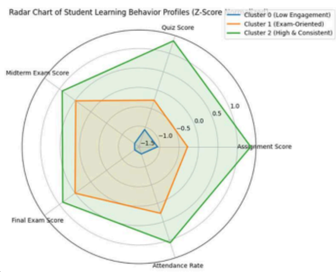

Figure 1. Radar Chart of Student Learning Behavior Profiles Across Three Clusters

The radar visualization clearly highlights distinct engagement patterns among clusters. Cluster 0 shows uniformly low values across all attributes, Cluster 2 displays consistently high and balanced engagement, and Cluster 1 demonstrates a distinctive profile with a stronger emphasis on examination performance relative to continuous assessment components. This visualization confirms that the identified clusters differ not only in overall engagement levels but also in the distribution of learning effort across academic activities.

To further emphasize the numerical differences between clusters, a centroid bar chart was used to visualize the mean values of each learning behavior attribute for each cluster, as shown in Figure 2.

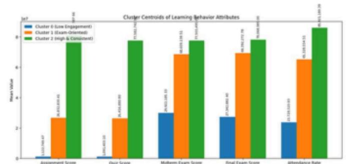

Figure 2. Cluster Centroids of Learning Behavior Attributes

The centroid bar chart reinforces the profiling results by showing clear and consistent differences in average values across clusters. The visualization confirms that engagement-related factors jointly contribute to cluster differentiation rather than isolated academic performance indicators.

The findings of this study demonstrate that student learning behavior over multiple semesters can be meaningfully categorized into three distinct patterns: low engagement, moderate exam-oriented engagement, and high consistent engagement. These patterns reflect differences not only in academic performance but also in learning strategies and levels of sustained participation.

The identification of an exam-oriented learning behavior pattern is particularly noteworthy. This pattern suggests that some students strategically allocate learning effort toward summative assessments while maintaining moderate engagement in continuous learning activities. This finding aligns with existing learning analytics literature, which highlights that summative-oriented evaluation systems may encourage strategic learning behaviors that previousitize examination performance over ongoing engagement.

Furthermore, the results indicate that learning behavior patterns represent longitudinal tendencies rather than short-term academic fluctuations. By utilizing data collected over eight semesters, this study provides evidence that stable learning behavior patterns can be identified using institutional academic data. These findings support theoretical distinctions between learning processes and learning outcomes and emphasize the importance of incorporating process-based indicators into academic evaluation systems.

CONCLUSION

This study investigated student learning behavior patterns using longitudinal academic data recorded in the Academic Information System (SIAKAD) through a clustering-based approach. By analyzing academic assessment components and attendance data collected over eight consecutive semesters, this research identified three distinct and stable learning behavior patterns among students. These patterns represent students with low learning engagement, students with moderate engagement characterized by an exam-oriented learning strategy, and students with high and consistent learning engagement across all learning components.

The findings demonstrate that student learning behavior is not homogeneous and cannot be adequately described using a binary classification of high and low performance. Instead, the results reveal meaningful variations in how students engage with learning activities over time. In particular, the identification of an exam-oriented learning behavior pattern highlights that some students strategically previousitize summative assessments while maintaining moderate levels of continuous engagement. This insight suggests that learning outcomes alone may not fully reflect the diversity of learning processes experienced by students throughout their academic journey.

From a practical perspective, the results of this study underscore the importance of incorporating process-based indicators, such as continuous assessment performance and attendance, into academic evaluation and monitoring systems. The ability to distinguish between low-engagement, exam-oriented, and highly engaged students provides valuable information for the development of targeted instructional strategies, academic advising, and early intervention mechanisms. By leveraging institutional academic data, higher education institutions can move toward more data-informed and responsive academic decision-making.

Despite its contributions, this study has several limitations. The analysis relied solely on academic assessment data and attendance records, which may not capture all dimensions of student learning behavior, such as motivation, learning preferences, or engagement in informal learning activities. Additionally, although the use of multi-semester data strengthens the stability of the identified patterns, the findings are based on data from a single institution and may not be directly generalizable to other educational contexts.

Future research may extend this study by incorporating additional data sources, such as Learning Management System (LMS) activity logs, survey-based engagement measures, or longitudinal qualitative data, to obtain a more comprehensive understanding of student learning behavior. Further studies may also explore the application of adaptive or hybrid clustering techniques and examine how learning behavior patterns evolve across different stages of students' academic trajectories.

REFERENCE

2_Similarity 113 Fauka

ORIGINALITY REPORT

3%	1%	2%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Central University of Punjab Student Paper	1%
2	Mohammad Abbas Fakharmanesh, Mojtaba Babaelahi. "Machine learning-driven metaheuristic optimization framework for enhanced stirling cryocooler performance: analysis of non-ideal heat transfer, pressure drop, and mechanical friction losses", Journal of Thermal Analysis and Calorimetry, 2025 Publication	1%
3	Submitted to De Montfort University Student Paper	<1%
4	B A Bulolo, M A Pase, F Ginting. "Antibiotic sensitivity pattern of bacteria from diabetic foot infections Haji Adam Malik central general hospital", IOP Conference Series: Earth and Environmental Science, 2018 Publication	<1%
5	Submitted to London School of Economics and Political Science Student Paper	<1%
6	Sylvia Frühwirth-Schnatter, Gilles Celeux, Christian P. Robert. "Handbook of Mixture Analysis", CRC Press, 2019 Publication	<1%