Vol. 8, No. 3, June 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Modeling Student Learning Profiles from LMS Behavioral Traces Using Big Data Analytics

Arief Hidayat[1], **Kusworo Adi**[2], **Bayu Surarso**[3]

[1]Informatics Department, Faculty of Engineering, Wahid Hasyim University, Semarang, Indonesia
[2,3]Informatics Department, Faculty of Mathematics and Sciences, Diponegoro University, Semarang, Indonesia

Received:
October 31, 2025
Revised:
April 11, 2026
Accepted:
May 30, 2026
Published:
June 22, 2026

Corresponding Author:
Author Name*:
Arief Hidayat
Email*:
arief.hidayat@unwahas.ac.id

DOI:
10.63158/journalisi.v8i3.1588

Abstract. Digital learning environments and Learning Management Systems (LMSs) generate large volumes of time-stamped behavioral traces that can be used to examine how students access resources, navigate course structures, communicate, and approach assessments. Traditional learning-style models often depend on static self-report categories and may not reflect how students actually study in digital courses. This study develops a learning analytics framework for modeling student learning profiles from authentic LMS behavioral traces. The study used a quantitative, non-experimental, longitudinal design based on Canvas LMS interaction data from 15,342 undergraduate students enrolled in 150 large-enrollment courses during the 2023–2024 academic year. More than 500 million raw interaction logs were processed into 24 engineered behavioral features representing temporal engagement, resource access, navigation behavior, interaction activity, and assessment timing. After feature normalization, K-Means clustering was applied, and the optimal cluster solution was selected using the elbow method and average silhouette score. Cluster distinctiveness was examined using one-way analysis of variance, and the association between cluster membership and academic performance category was evaluated using a Chi-squared test. The analysis supported a four-cluster solution. Assessment procrastination and navigation sequentially were the strongest differentiating features.

Keywords: Learning Analytics, Educational Data Mining, LMS Behavioral Traces, K-Means Clustering, Student Behavioral Modeling, Learning Profiles

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Digital learning environments have changed how higher education records student activity. Learning Management Systems (LMSs), Massive Open Online Courses, and other online learning platforms routinely capture time-stamped traces of resource access, navigation, discussion activity, video use, and assessment behavior [1]–[4]. These traces allow researchers and instructors to study learning behavior as it occurs in authentic digital environments rather than relying only on surveys or end-of-course performance data.

Learning analytics and educational data mining have developed as complementary fields for transforming educational traces into interpretable evidence for learning design, student support, and institutional decision-making [1]–[5]. In large-enrollment courses, instructors often need to understand differences in pacing, planning, resource use, interaction, and assessment behavior, but manual diagnosis is difficult. LMS data can support this need because they provide scalable indicators of what students do across a course, when they do it, and how their behaviors relate to learning outcomes [6]–[9].

Traditional learning-style models, such as VARK, Kolb's Learning Cycle, and the Felder-Silverman model, have often been used to explain individual differences in learning. However, these models rely heavily on self-report instruments and fixed categories. Reviews of learning-style research have questioned whether matching instruction to declared learning styles improves learning outcomes and have warned against treating such categories as stable learner traits [10]–[12]. Therefore, a more defensible approach is to study observable learning behavior and describe patterns as dynamic profiles rather than permanent types.

This behavioral perspective aligns with self-regulated learning research. Self-regulated learning emphasizes how students plan, monitor, control, and adapt their learning strategies across time and context [13]–[15]. LMS traces do not capture every cognitive or motivational process, but they can provide observable indicators of planning, time management, navigation, help-seeking, and assessment preparation. For example, early access to assessment materials may indicate planning, while late clustered access may indicate deadline-driven engagement.

Previous learning analytics studies have used LMS data to identify engagement patterns, predict achievement, and detect at-risk students [6]–[9], [16], [17]. However, many studies rely on a small number of indicators, such as login frequency, time-on-task, or quiz scores. Fewer studies construct multivariate behavioral profiles that combine temporal, navigational, resource-use, interaction, and assessment-related dimensions in one framework. This gap is important because student engagement in an LMS is not a single behavior but a configuration of actions across the digital learning environment.

This study addresses that gap by developing a learning analytics framework for modeling student learning profiles from LMS behavioral traces. The framework does not assign students to predefined learning-style categories. Instead, it uses engineered behavioral features and unsupervised clustering to identify naturally occurring patterns of digital engagement. The study was guided by three research questions: RQ1: What behavioral learning profiles emerge from large-scale LMS interaction data? RQ2: Which engineered behavioral features most strongly differentiate these profiles? RQ3: Are the identified behavioral profiles significantly associated with academic performance categories?

The expected contribution is both methodological and practical. Methodologically, the study reframes learning style as an emergent behavioral profile that can be measured from digital traces. Practically, the resulting profiles can support adaptive learning systems, instructor dashboards, and targeted student support when interpreted as context-sensitive patterns rather than deterministic labels.

2. METHODS

2.1. Research Design

This study used a quantitative, non-experimental, longitudinal design within the fields of learning analytics and educational data mining. The design was descriptive and correlational. First, LMS log data were transformed into student-level behavioral features. Second, unsupervised clustering was used to identify learning profiles. Third, statistical tests evaluated whether the profiles differed on key engineered features and whether cluster membership was associated with final academic performance category.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The research workflow consisted of data extraction, data cleaning, anonymization, student-course aggregation, sessionization, feature engineering, normalization, K-Means clustering, cluster validation, statistical testing, and pedagogical interpretation. This workflow is summarized in Figure 1.

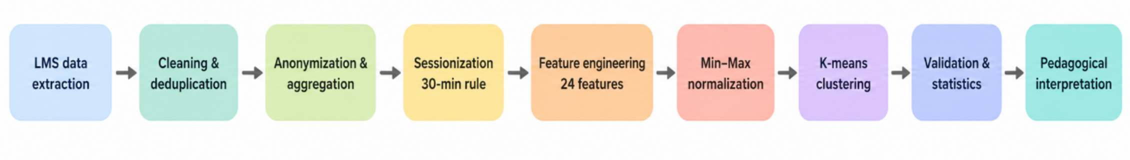

Figure 1. Research workflow for modeling LMS-derived behavioral learning profiles

All procedures involving student data followed institutional ethical requirements. The research protocol received approval from the university's Institutional Review Board before data access. The dataset was anonymized at the source by removing personally identifiable information and replacing student identifiers with non-reversible hashed IDs. The analysis was conducted at aggregate behavioral-profile level and was not used for punitive individual student evaluation.

2.2. Population and Sample

The target population comprised undergraduate students enrolled at a large, multi-disciplinary, public research university in North America. The institution used a centralized Canvas LMS, which provided a consistent source of interaction data across academic disciplines. The initial sampling frame included approximately 22,000 students active during the 2023–2024 academic year. Data were drawn from 150 large-enrollment introductory courses, defined as courses with more than 100 enrolled students, across STEM, humanities, and social sciences. Students with fewer than 100 total LMS log entries and students who officially withdrew before the census date were excluded. The final analyzable sample consisted of 15,342 unique students and more than 500 million time-stamped LMS interaction logs.

2.3. Data Source and Variables

The primary data source was the university's Canvas LMS. The platform automatically recorded user-initiated interactions, including page views, resource access, module navigation, discussion activity, video player events, assessment interactions, timestamps, and course performance indicators. The raw dataset contained four categories of variables: (1) clickstream and navigation data, including resource views, page loads, and module transitions; (2) temporal data, including login timestamps, session timing, and time-of-day access; (3) interaction data, including forum posts, replies, peer-review actions, and video player events; and (4) academic performance data, including quiz, assignment, and final course grade records.

2.4. Data Processing and Feature Engineering

Data were extracted from the protected institutional data warehouse using secure SQL queries. Preprocessing included removal of erroneous records, deduplication of repeated log events, anonymized student-course aggregation, and sessionization of time-stamped events into learning sessions. A new learning session was defined after 30 minutes of inactivity between consecutive LMS events. Sessions shorter than 10 seconds and duplicate events with the same student, course, event type, resource identifier, and timestamp were removed. Extremely large feature values were winsorized at the 1st and 99th percentiles before normalization to reduce the influence of outliers while preserving the student-level record.

Raw interactions were converted into 24 engineered student-level behavioral features. Features were first computed at the student-course level and then aggregated to a unique student-level record across the included courses. The 24 features represented temporal engagement, resource use, navigation behavior, social interaction, assessment activity, monitoring behavior, and assessment timing. Table 1 reports descriptive statistics for six salient features, and Table 2 provides the complete feature dictionary used for clustering.

Table 1. Descriptive statistics of key engineered features (N = 15,342).

Feature ID	Feature Description	Mean	SD	Min	Max
F1	Mean session duration (minutes)	28.7	14.2	3.1	92.5
F2	PDF/Text access frequency	65.4	22.1	5.0	180.0
F3	Video/Media access frequency	45.2	30.5	0.0	150.0
F4	Assessment procrastination index (0–1)	0.68	0.21	0.05	1.00
F5	Navigation sequentially score (0–1)	0.55	0.19	0.10	0.95
F6	Forum contribution count	8.3	7.1	0.0	45.0

Table 2. Engineered behavioral feature dictionary used for clustering.

Feature ID	Feature Name	Dimension	Operational Definition	Formula or Computation Rule	Rationale
F1	Mean session duration	Temporal engagement	Average duration of valid LMS sessions in minutes.	Σ(session_end − session_start) / number of valid sessions; new sessions begin after 30 minutes of inactivity.	Depth of time-based engagement.
F2	PDF/Text access frequency	Resource use	Number of text-based resource views or downloads.	Count of unique PDF, page, and text-resource access events per student after duplicate removal.	Use of text-based learning materials.
F3	Video/Media access frequency	Resource use	Number of video or media access events.	Count of video play, replay, pause, completion, and media page events per student.	Multimedia learning behavior.
F4	Assessment procrastination index	Assessment timing	Relative timing of assessment activity in relation to release time and deadline.	Mean[1 − ((deadline − first_assessment_event) / (deadline − release_time))], bounded to 0–1; higher values indicate later activity.	Deadline-driven assessment behavior.
F5	Navigation sequentiality score	Navigation behavior	Degree to which module transitions followed the instructor-defined sequence.	Number of adjacent module transitions matching the designed next item / total module transitions.	Linear versus exploratory navigation.
F6	Forum contribution count	Interaction behavior	Number of discussion posts and replies.	Count of discussion thread posts, replies, and peer-response events per student.	Peer interaction and help-seeking.

Feature ID	Feature Name	Dimension	Operational Definition	Formula or Computation Rule	Rationale
F7	Login frequency	Temporal engagement	Number of LMS login or course-entry events.	Count of authenticated course-entry events per student across included courses.	Frequency of platform engagement.
F8	Active days ratio	Temporal engagement	Proportion of course days with at least one LMS event.	Number of active LMS days / number of enrolled course days.	Continuity of engagement across the semester.
F9	Evening/weekend access ratio	Temporal engagement	Proportion of LMS events occurring after 18:00 or on weekends.	Evening and weekend events / total LMS events.	Temporal preference and workload timing.
F10	Session regularity score	Temporal engagement	Consistency of study-session spacing across the semester.	1 − normalized coefficient of variation of inter-session intervals.	Regular versus bursty study rhythm.
F11	Assignment page view count	Assessment activity	Number of assignment-page views.	Count of assignment-instruction and assignment-submission page views.	Attention to assessment requirements.
F12	Quiz preparation event count	Assessment activity	Number of pre-quiz review or quiz-information events.	Count of quiz instructions, practice quiz, rubric, and related resource events before first quiz attempt.	Assessment preparation behavior.
F13	Deadline buffer hours	Assessment timing	Average time between first	Mean(deadline − first_assessment_event) in hours for graded assessments.	Lead time before deadlines.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Feature ID	Feature Name	Dimension	Operational Definition	Formula or Computation Rule	Rationale
			assessment-related activity and deadline.		
F14	Module completion rate	Navigation behavior	Proportion of instructor-defined module items accessed.	Number of required module items accessed / number of available required items.	Breadth of course-material coverage.
F15	Backtracking/nonlinear transition rate	Navigation behavior	Proportion of module transitions that moved backward or skipped the designed sequence.	Non-sequential transitions / total module transitions.	Exploratory or fragmented navigation.
F16	Search event count	Navigation behavior	Number of LMS search events.	Count of internal search and resource-filter events per student.	Problem-focused navigation.
F17	Announcement view count	Interaction behavior	Number of course announcement views.	Count of announcement page views or notification clicks.	Monitoring of instructor communications.
F18	Peer reply count	Interaction behavior	Number of replies to peers in discussion spaces.	Count of replies to student-originated posts.	Peer-to-peer academic interaction.
F19	Instructor message count	Interaction behavior	Number of direct LMS messages or instructor-contact events.	Count of inbox messages, instructor replies, or help-contact events recorded in the LMS.	Help-seeking from teaching staff.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Feature ID	Feature Name	Dimension	Operational Definition	Formula or Computation Rule	Rationale
F20	Video completion ratio	Resource use	Proportion of video events reaching completion threshold.	Video completion events / video play events; completion threshold set at 80% watched.	Depth of media use.
F21	Average video pause/replay count	Resource use	Average pause and replay events per watched video.	(Pause events + replay/seek-back events) / number of videos accessed.	Review-oriented video behavior.
F22	Early resource access ratio	Temporal/resource use	Proportion of resource events occurring before the final 25% of the assessment window.	Early resource events / total resource events linked to assessment windows.	Early preparation rather than last-minute access.
F23	Assessment attempt count	Assessment activity	Number of graded assessment attempts or submissions.	Count of quiz attempts, assignment submissions, and resubmissions where allowed.	Assessment activity intensity.
F24	Gradebook checking frequency	Monitoring behavior	Number of gradebook or feedback views.	Count of gradebook, rubric feedback, and score-report views.	Performance monitoring behavior.

2.5. Clustering Procedure

All clustering features were normalized using Min-Max scaling before modeling so that variables with larger numeric ranges did not dominate distance calculations. K-Means clustering was selected because it is a widely used partitioning method for grouping observations by feature-space similarity [18], [19]. Candidate solutions from k = 2 to k = 10 were compared using the elbow method and average silhouette score [20].

Clustering was implemented in Python 3.10 using scikit-learn KMeans with Euclidean distance, k-means++ initialization, n_init = 20 random starts, max_iter = 300, convergence tolerance = 1×10^{-4}, and random_state = 42. The final cluster solution was selected based on the combined evidence of within-cluster compactness, between-cluster separation, interpretability, and parsimony. The k = 4 solution was retained because it showed the clearest elbow, the highest average silhouette score, and interpretable behavioral differences across the resulting clusters.

2.6. Statistical Analysis

After clustering, one-way analysis of variance (ANOVA) tested whether the clusters differed significantly across the engineered behavioral features. Tukey's HSD was used for post-hoc pairwise comparisons where appropriate. Effect size was reported using eta-squared (η^2). A Chi-squared test of independence examined the relationship between cluster membership and academic performance categories: High (A), Medium (B/C), and Low (D/F/Withdraw). The effect size for the Chi-squared test was reported using Cramer's V. Because the study was observational, all statistical relationships were interpreted as associations rather than causal effects.

3. RESULTS AND DISCUSSION

The final preprocessed dataset consisted of 15,342 students and more than 500 million raw LMS interaction logs. Each student was represented by 24 engineered behavioral features covering temporal, navigational, resource-use, interaction, monitoring, and assessment-related dimensions. Descriptive statistics for six salient engineered features are shown in Table 1. Min-Max scaling was applied before clustering, but the table reports pre-normalized values to show the original behavioral variation in the sample. The descriptive statistics indicate substantial variability in LMS behavior. Mean session duration had a standard deviation of 14.2 minutes, suggesting variation between short check-ins and longer study sessions. The assessment procrastination index had a mean of 0.68 and a standard deviation of 0.21, indicating that many students engaged with assessments close to deadlines while others began earlier.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Resource-use patterns also varied across students. PDF/Text access frequency had a higher mean than Video/Media access frequency, but video use showed greater dispersion. This pattern suggests that text resources were broadly used, whereas video resources were more unevenly distributed across the sample. The navigation sequentially scores had a mean of 0.55, indicating that students combined linear module progression with exploratory navigation.

3.1. Cluster Selection

Cluster selection was conducted by comparing k values from 2 to 10. The elbow method showed a clear bend at k = 4, after which additional clusters produced smaller reductions in within-cluster sum of squares. The average silhouette score also supported k = 4, with the highest observed value of 0.42. The final analysis therefore used a four-cluster solution.

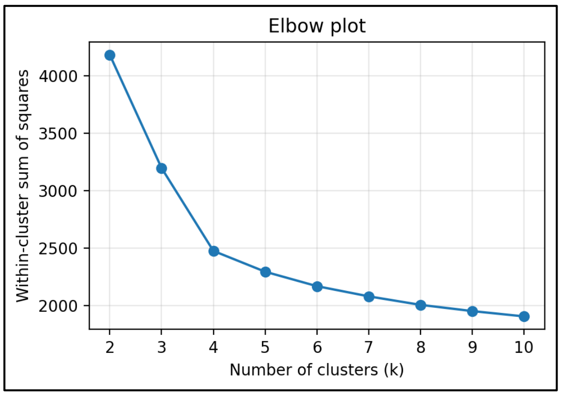

Figure 2. Elbow plot for candidate K-Means solutions from k = 2 to k = 10.

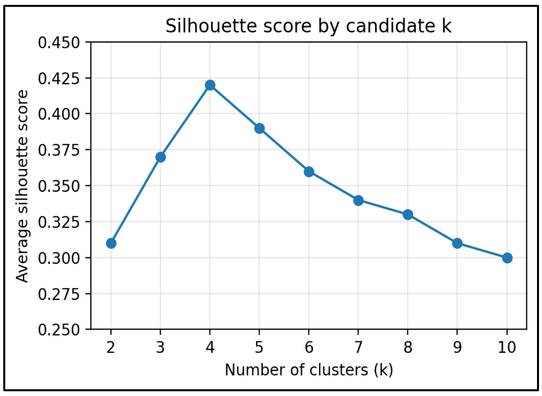

Figure 3. Average silhouette scores for candidate K-Means solutions from k = 2 to k = 10.

The resulting clusters were unequal in size: C1 accounted for 28.5% of the sample, C2 for 35.1%, C3 for 19.8%, and C4 for 16.6%. This distribution indicates that the student population was behaviorally heterogeneous rather than concentrated in one dominant engagement pattern.

Table 3. Cluster size distribution.

Cluster	Profile Label	N	Percentage
C1	Sequential Planners	4,372	28.5%
C2	Regular Resource Balancers	5,384	35.1%

Cluster	Profile Label	N	Percentage
C3	Exploratory Multimedia Users	3,039	19.8%
C4	Deadline-driven Navigators	2,547	16.6%
	Total	15,342	100.0%

3.2. Cluster Differences Across Engineered Features

ANOVA results showed statistically significant differences among the four clusters across the engineered features (all $p < .001$). The largest differences were observed for assessment procrastination index, $F(3, 15338) = 874.2$, $p < .001$, $\eta^2 = .146$, and navigation sequentially score, $F(3, 15338) = 712.9$, $p < .001$, $\eta^2 = .122$. Tukey's HSD comparisons indicated significant pairwise differences among clusters on these two dimensions.

Table 4. Complete cluster-profile table for the 24 engineered features.

Feature	C1 M (SD)	C2 M (SD)	C3 M (SD)	C4 M (SD)	ANOVA p	η^2
F1 Mean session duration	32.10 (13.80)	29.00 (13.50)	26.50 (14.00)	24.80 (12.70)	< .001	.036
F2 PDF/Text access frequency	85.40 (20.40)	70.00 (18.90)	58.10 (21.10)	30.10 (17.60)	< .001	.469
F3 Video/Media access frequency	30.00 (21.80)	38.00 (24.50)	51.60 (27.40)	78.90 (31.20)	< .001	.300
F4 Assessment procrastination index	0.31 (0.10)	0.80 (0.16)	0.81 (0.14)	0.91 (0.05)	< .001	.146
F5 Navigation sequentially score	0.82 (0.09)	0.55 (0.15)	0.42 (0.16)	0.24 (0.08)	< .001	.122
F6 Forum contribution count	5.00 (4.20)	7.00 (5.30)	9.60 (6.40)	15.20 (8.10)	< .001	.261
F7 Login frequency	72.00 (23.00)	64.00 (21.00)	52.00 (19.00)	58.00 (25.00)	< .001	.097
F8 Active days ratio	0.61 (0.14)	0.53 (0.16)	0.43 (0.17)	0.38 (0.15)	< .001	.227
F9 Evening/weekend access ratio	0.32 (0.18)	0.42 (0.20)	0.54 (0.21)	0.68 (0.19)	< .001	.289
F10 Session regularity score	0.76 (0.12)	0.58 (0.15)	0.42 (0.17)	0.24 (0.13)	< .001	.610
F11 Assignment page view count	42.00 (14.00)	38.00 (13.00)	31.00 (12.00)	24.00 (11.00)	< .001	.197
F12 Quiz preparation event count	10.00 (4.50)	8.00 (4.00)	6.50 (3.80)	4.00 (3.20)	< .001	.204
F13 Deadline buffer hours	64.00 (22.00)	30.00 (18.00)	18.00 (14.00)	6.00 (6.00)	< .001	.599
F14 Module completion rate	0.91 (0.08)	0.78 (0.12)	0.65 (0.16)	0.51 (0.18)	< .001	.525
F15 Backtracking/nonlinear transition rate	0.18 (0.09)	0.32 (0.13)	0.48 (0.16)	0.66 (0.14)	< .001	.624
F16 Search event count	6.00 (5.00)	9.00 (6.00)	14.00 (8.00)	18.00 (9.00)	< .001	.285
F17 Announcement view count	28.00 (9.00)	24.00 (8.50)	19.00 (8.00)	16.00 (7.50)	< .001	.210
F18 Peer reply count	3.80 (3.20)	5.00 (4.10)	7.50 (5.20)	12.00 (6.70)	< .001	.271
F19 Instructor message count	1.20 (1.40)	1.00 (1.20)	1.30 (1.50)	1.90 (1.80)	< .001	.043
F20 Video completion ratio	0.58 (0.20)	0.64 (0.18)	0.71 (0.17)	0.76 (0.15)	< .001	.113
F21 Average video pause/replay count	1.80 (1.50)	2.40 (1.70)	3.10 (2.00)	4.20 (2.30)	< .001	.167
F22 Early resource access ratio	0.70 (0.17)	0.42 (0.18)	0.30 (0.16)	0.12 (0.10)	< .001	.604
F23 Assessment attempt count	1.30 (0.60)	1.50 (0.70)	1.80 (0.90)	2.40 (1.10)	< .001	.181
F24 Gradebook checking frequency	14.00 (6.00)	12.00 (5.50)	9.50 (5.00)	8.50 (4.80)	< .001	.123

The cluster profiles show four interpretable behavioral patterns. Cluster 1, labeled Sequential Planners, had the lowest assessment procrastination index ($M = 0.31$, $SD = 0.10$), the highest navigation sequentially scores ($M = 0.82$, $SD = 0.09$), high text-resource use,

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

high module completion, strong session regularity, and the largest deadline buffer. This profile represents students who tended to follow the instructor-defined instructional pathway and begin assessment-related work earlier. Cluster 2, labeled Regular Resource Balancers, was the largest group. Students in this profile showed moderate session duration, high text-resource access, moderate navigation sequentially, and moderate interaction levels. Their behavior suggests regular but less strongly planned engagement than C1, with balanced use of course materials and assessment pages. Cluster 3, labeled Exploratory Multimedia Users, showed lower sequentially than C2, higher video/media access, higher search activity, and more peer replies. This profile suggests a more exploratory pattern in which students used multimedia resources and peer interaction to support learning while moving less linearly through the LMS structure. Cluster 4, labeled Deadline-driven Navigators, had the highest assessment procrastination index (M = 0.91, SD = 0.05), low navigation sequentially (M = 0.24, SD = 0.08), low text-resource use, high video/media access, high forum participation, and the lowest deadline buffer. This profile indicates a just-in-time and non-linear learning strategy rather than simple disengagement.

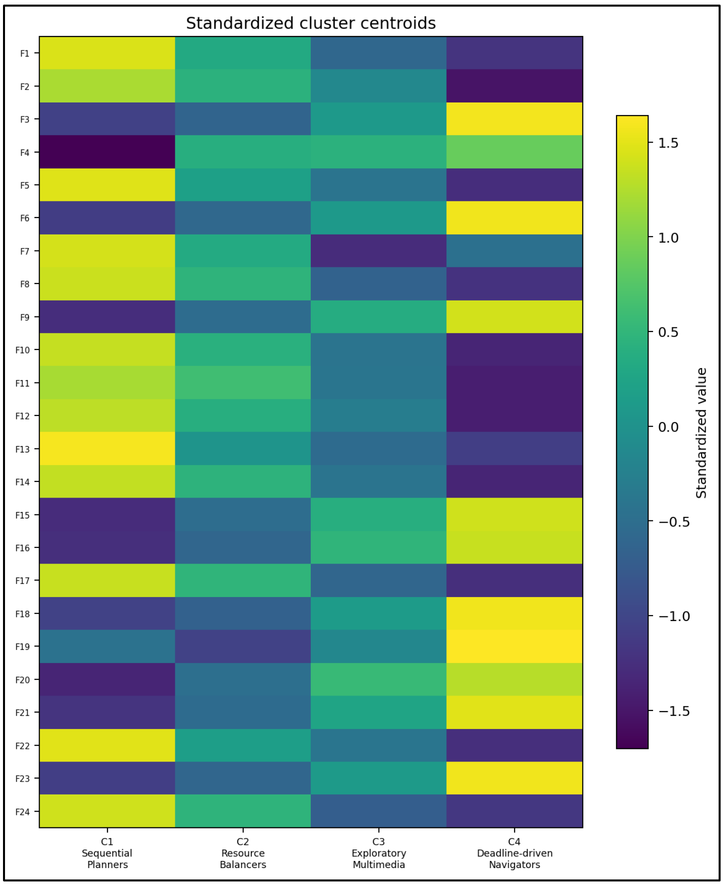

Figure 4. Standardized centroid heatmap for the four-cluster solution.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.3. Association Between Cluster Membership and Academic Performance

The pedagogical relevance of the clusters was examined using academic performance categories. Final course grades were grouped as High (A), Medium (B/C), and Low (D/F/Withdraw). The Chi-squared test showed a statistically significant association between cluster membership and academic performance, $\chi^2(6, N = 15{,}342) = 1045.7$, $p < .001$. The corresponding effect size was Cramer's V = 0.185, indicating a small-to-moderate association. This result indicates that the behavioral profiles were related to course outcomes, although the observational design does not support causal claims.

Table 5. Cluster membership by academic performance category.

Cluster	High (A)	Medium (B/C)	Low (D/F/Withdraw)	Total
C1 Sequential Planners	1,704	1,904	764	4,372
C2 Regular Resource Balancers	1,708	2,329	1,347	5,384
C3 Exploratory Multimedia Users	726	1,728	585	3,039
C4 Deadline-driven Navigators	332	1,078	1,137	2,547
Total	4,470	7,039	3,833	15,342

3.4. Discussion

This study shows that LMS behavioral traces can be used to model meaningful student learning profiles. The four-cluster solution indicates that students differ not only in how often they access the LMS but also in when they engage, how they navigate course structures, which resources they use, how they interact, and how they approach assessments. The findings extend learning analytics and educational data mining research by moving beyond isolated engagement indicators toward multivariate behavioral profiles [1]–[5]. Instead of treating login frequency or time-on-task as sufficient indicators of learning, this study combines temporal, resource-use, interaction, navigation, monitoring, and assessment features into a broader description of digital learning behavior.

The strongest differentiators were assessment procrastination and navigation sequentially. This suggests that students' relationships to deadlines and course structure are central to how they organize learning in the LMS. Sequential Planners appeared to follow the intended instructional path and begin assessment-related work earlier, whereas Deadline-driven Navigators used a more reactive and non-linear strategy.

The four profiles should not be interpreted as permanent student types. They represent behavioral strategies enacted within a specific technological, institutional, and course-design context. A student's cluster membership may change across courses, semesters, assessment designs, or instructional formats. The profiles are therefore best understood as context-sensitive learning behaviors rather than fixed learning styles. This distinction is important because the learning-style literature has been criticized for relying on stable self-report categories that are not consistently supported by experimental evidence [10]–[12].

The findings also align with self-regulated learning theory. Assessment timing may reflect planning and time management, while navigation sequentially may reflect how students monitor and organize their progression through course materials [13]–[15]. However, LMS traces should be interpreted cautiously. Log data record observable actions, but they do not directly reveal motivation, cognition, comprehension, or learning intention. The practical value of the framework lies in formative support. Instructors could use cluster-level information to identify whether a class is dominated by sequential, balanced, exploratory, or deadline-driven patterns. Student-facing dashboards could also present private, supportive feedback about time management, navigation strategies, and resource-use habits. Such dashboards should help students interpret their behavior and plan improvement rather than label them rigidly [21], [25], [26].

The Deadline-driven Navigator profile should not be reduced to disengagement. Its higher use of video resources and forums suggests a just-in-time learning strategy supported by digital platform affordances. Nevertheless, this strategy may carry risks when students rely on rapid, fragmented resource access without sufficient time for conceptual consolidation. Instructors can respond through milestone reminders, low-stakes checkpoints, clearer links between assessments and supporting resources, and structured alternative pathways.

The study has several limitations. First, the data came from one institution, so the generalizability of the four-cluster solution requires replication in other institutional, disciplinary, and course-design contexts. Second, the analysis assigned students to dominant semester-level profiles and did not model week-to-week transitions. Third, LMS traces capture platform-visible behavior but not offline study, private collaboration, motivation, or comprehension. Fourth, although the expanded feature table improves reproducibility, the behavioral meaning of each feature may vary across course designs and LMS configurations.

Future research should replicate the framework across institutional types, disciplines, and course modalities. Temporal models, such as sequence analysis or Hidden Markov Models, could examine how students move between behavioral profiles over time. Intervention studies could then test whether targeted nudges, adaptive resource recommendations, or redesigned assessment schedules improve engagement and academic outcomes.

4. CONCLUSION

This study demonstrates that large-scale LMS behavioral traces can support the modeling of student learning profiles using K-Means clustering. Using data from 15,342 students and more than 500 million interaction logs, the analysis identified a four-cluster solution that differentiated students by temporal engagement, resource use, navigation behavior, forum participation, monitoring behavior, and assessment timing. The four profiles were labeled Sequential Planners, Regular Resource Balancers, Exploratory Multimedia Users, and Deadline-driven Navigators. Assessment procrastination and navigation sequentially were the strongest differentiating features. Cluster membership was significantly associated with academic performance categories, $\chi^2(6, N = 15{,}342) = 1045.7$, $p < .001$, Cramer's $V = .185$, but this association should not be interpreted as causal. The study contributes a behavior-first alternative to static, self-reported learning-style approaches by reframing learning style as an emergent and context-sensitive pattern of digital engagement. Future work should replicate the framework across institutions and apply temporal modeling to capture changes in student behavior across courses and semesters.

ACKNOWLEDGMENT

The authors thank the Software Engineering Laboratory at Wahid Hasyim University for providing facilities, software resources, and a supportive research environment. The authors also acknowledge the institutional data governance and ethics procedures that enabled anonymized analysis of LMS interaction data.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

REFERENCES

[1] G. Siemens, "Learning analytics: The emergence of a discipline," American Behavioral Scientist, vol. 57, no. 10, pp. 1380–1400, 2013, doi: 10.1177/0002764213498851.

[2] R. Ferguson, "Learning analytics: drivers, developments and challenges," International Journal of Technology Enhanced Learning, vol. 4, no. 5/6, pp. 304–317, 2012, doi: 10.1504/IJTEL.2012.051816.

[3] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," WIREs Data Mining and Knowledge Discovery, vol. 10, no. 3, e1355, 2020, doi: 10.1002/widm.1355.

[4] R. S. J. d. Baker and G. Siemens, "Educational data mining and learning analytics," in The Cambridge Handbook of the Learning Sciences, 2nd ed., R. K. Sawyer, Ed. Cambridge: Cambridge University Press, 2014, pp. 253–272.

[5] D. Gašević, S. Dawson, and G. Siemens, "Let's not forget: Learning analytics are about learning," TechTrends, vol. 59, no. 1, pp. 64–71, 2015, doi: 10.1007/s11528-014-0822-x.

[6] O. Viberg, M. Hatakka, O. Bälter, and A. Mavroudi, "The current landscape of learning analytics in higher education," Computers in Human Behavior, vol. 89, pp. 98–110, 2018, doi: 10.1016/j.chb.2018.07.027.

[7] R. Cerezo, M. Sánchez-Santillán, M. P. Paule-Ruiz, and J. C. Núñez, "Students' LMS interaction patterns and their relationship with achievement: A case study in higher education," Computers & Education, vol. 96, pp. 42–54, 2016, doi: 10.1016/j.compedu.2016.02.006.

[8] J. W. You, "Identifying significant indicators using LMS data to predict course achievement in online learning," The Internet and Higher Education, vol. 29, pp. 23–30, 2016, doi: 10.1016/j.iheduc.2015.11.003.

[9] J. Jovanović, D. Gašević, S. Dawson, A. Pardo, and N. Mirriahi, "Learning analytics to unveil learning strategies in a flipped classroom," The Internet and Higher Education, vol. 33, pp. 74–85, 2017, doi: 10.1016/j.iheduc.2017.02.001.

[10] H. Pashler, M. McDaniel, D. Rohrer, and R. Bjork, "Learning styles: Concepts and evidence," Psychological Science in the Public Interest, vol. 9, no. 3, pp. 105–119, 2008, doi: 10.1111/j.1539-6053.2009.01038.x.

[11] P. A. Kirschner, "Stop propagating the learning styles myth," Computers & Education, vol. 106, pp. 166–171, 2017, doi: 10.1016/j.compedu.2016.12.006.

[12] P. M. Newton and M. Miah, "Evidence-based higher education: Is the learning styles myth important?" Frontiers in Psychology, vol. 8, article 444, 2017, doi: 10.3389/fpsyg.2017.00444.

[13] P. H. Winne and A. F. Hadwin, "Studying as self-regulated learning," in Metacognition in Educational Theory and Practice, D. J. Hacker, J. Dunlosky, and A. C. Graesser, Eds. Mahwah, NJ: Lawrence Erlbaum, 1998, pp. 277–304.

[14] E. Panadero, "A review of self-regulated learning: Six models and four directions for research," Frontiers in Psychology, vol. 8, article 422, 2017, doi: 10.3389/fpsyg.2017.00422.

[15] J. Broadbent and W. L. Poon, "Self-regulated learning strategies and academic achievement in online higher education learning environments: A systematic review," The Internet and Higher Education, vol. 27, pp. 1–13, 2015, doi: 10.1016/j.iheduc.2015.04.007.

[16] C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," Computers & Education, vol. 131, pp. 22–32, 2019, doi: 10.1016/j.compedu.2018.12.006.

[17] L. A. Macfadyen and S. Dawson, "Mining LMS data to develop an early warning system for educators: A proof of concept," Computers & Education, vol. 54, no. 2, pp. 588–599, 2010, doi: 10.1016/j.compedu.2009.09.008.

[18] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, Berkeley, CA: University of California Press, 1967, pp. 281–297.

[19] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2007, pp. 1027–1035.

[20] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[21] S. Slade and P. Prinsloo, "Learning analytics: Ethical issues and dilemmas," American Behavioral Scientist, vol. 57, no. 10, pp. 1510–1529, 2013, doi: 10.1177/0002764213479366.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[22] D. Ifenthaler and C. Schumacher, "Student perceptions of privacy principles for learning analytics," Educational Technology Research and Development, vol. 64, pp. 923–938, 2016, doi: 10.1007/s11423-016-9477-y.

[23] T. Susnjak, G. S. Ramaswami, and A. Mathrani, "Learning analytics dashboard: A tool for providing actionable insights to learners," International Journal of Educational Technology in Higher Education, vol. 19, article 12, 2022, doi: 10.1186/s41239-021-00313-7.

[24] M. Blumenstein, "Synergies of learning analytics and learning design: A systematic review of student outcomes," Journal of Learning Analytics, vol. 7, no. 3, pp. 13–32, 2020, doi: 10.18608/jla.2020.73.3.

[25] L. Paulsen and E. Lindsay, "Learning analytics dashboards are increasingly becoming about learning and not just analytics: A systematic review," Education and Information Technologies, vol. 29, no. 11, pp. 14279–14308, 2024, doi: 10.1007/s10639-023-12401-4.

[26] I. Molenaar, "Towards hybrid human-AI learning technologies," European Journal of Education, vol. 57, no. 4, pp. 632–645, 2022, doi: 10.1111/ejed.12527.

[27] S. H. E. Dijkstra, M. Hinne, E. Segers, and I. Molenaar, "Clustering children's learning behaviour to identify self-regulated learning support needs," Computers in Human Behavior, vol. 145, article 107754, 2023, doi: 10.1016/j.chb.2023.107754.