International Journal of Nexural Intelligence
Vol. 1, No. 1, Mei 2026

Learning Analytics-Based Student Engagement Profiling Using Moodle Activity Logs Across Multiple Courses

Abdul Rahman[1], Destiarini[2]
[1,2] Informatics Study Program, Faculty of Engineering and Computer Science, Universitas Baturaja, South Sumatera, Indonesia

Article Info

Article history:

Received April 17, 2026
Accepted Mei 19, 2026
Available Mei 24, 2026

Keywords:

Learning analytics
Moodle logs
Student engagement
Educational data mining
Clustering

ABSTRACT

Learning management systems generate detailed event logs, yet these data are often underused for monitoring student engagement. This study develops a learning analytics workflow for profiling engagement from Moodle activity logs collected from 20 courses. The raw dataset contained 402,290 records recorded between 19 February 2025 and 2 June 2025. After removing administrative, reporting, and system-maintenance events, 369,592 learner-generated events from 760 users were analyzed. The method consisted of timestamp parsing, course mapping, anonymization, event-category mapping, feature engineering, engagement scoring, and K-Means clustering. Eight behavioral indicators were constructed, including total events, active days, course views, module views, quiz activity, assignment activity, resource access, and event diversity. The results show that quiz-related interactions dominated the logs, followed by system course views and assignment activities. Three engagement profiles were identified: low, moderate, and high engagement. The proposed workflow provides an interpretable basis for course monitoring and early identification of learners who may require academic support, while avoiding unsupported claims about academic achievement when final-grade data are unavailable.

Corresponding Author:

Abdul Rahman
Informatics Study Program, Faculty of Engineering and Computer Science, Universitas Baturaja
Ki Ratu Penghulu No. 2301, Tanjung Baru, Subdistrict Baturaja Timur, Ogan Komering Ulu Regency, South Sumatera, Indonesia
Email: abdulrahman@unbara.ac.id

1. INTRODUCTION

Digital learning environments have become important institutional infrastructures for recording how learners interact with course materials, assessments, and communication tools. Moodle, as one of the most widely used learning management systems, records detailed event logs that can describe what learners access, when they access it, and how frequently they interact with learning activities. These event traces are valuable for learning analytics because they provide observable behavioral indicators that are difficult to capture through conventional attendance or end-of-semester reports [1], [2].

Learning analytics and educational data mining have been used to transform raw educational data into actionable information for teachers, academic managers, and student-support units [3], [4]. Previous studies indicate that log-based indicators such as clicks, access frequency, active days, assessment interaction, resource viewing, and online participation are commonly used as proxies for behavioral engagement [2], [5], [6]. However, raw Moodle logs are not immediately ready for interpretation because they contain administrative actions, system-generated records, duplicated event semantics, and context-specific activity names.

The research problem addressed in this study is the gap between the availability of large Moodle event logs and the need for simple, interpretable engagement profiles at course level. In many institutions, logs are exported mainly for administrative auditing, while their analytical potential for academic monitoring remains limited [7]. Without systematic preprocessing and feature construction, lecturers may find it difficult to identify which learners interact consistently and which learners show limited participation.

This study proposes a descriptive and unsupervised learning analytics workflow for profiling student engagement using Moodle log files from multiple courses. The workflow includes data cleaning, anonymization, event-category mapping, feature engineering, engagement scoring, and clustering. The contribution of this study is practical and methodological: it shows how institutional Moodle logs can be converted into structured engagement indicators without requiring final-grade data. Therefore, the study focuses on engagement profiling, not academic performance prediction.

2. METHOD

This study used a quantitative learning analytics design based on secondary Moodle event logs. The research procedure followed six stages: data acquisition, data cleaning, anonymization, event-category mapping, feature engineering, and engagement profiling. Figure 1 summarizes the workflow applied in this study.

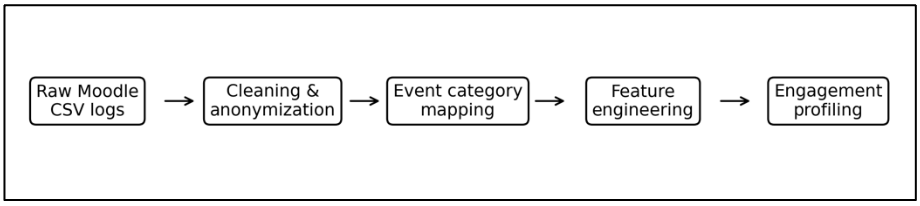

Figure 1. Shows the flowchart of the AI-based models and experimental methods applied

2.1. Dataset and preprocessing

The dataset consisted of 20 comma-separated value files exported from Moodle course logs. The combined raw data contained 402,290 records with nine fields: Time, User full name, Affected user, Event context, Component, Event name, Description, Origin, and IP address. The logs covered the period from 19 February 2025 to 2 June 2025. Because the exported data contained personally identifiable information, participant names were not used for reporting. Instead, actor identifiers extracted from the Description field were used only for aggregation and were treated as anonymized internal identifiers.

Data preprocessing removed administrative, log-reporting, enrollment-management, grading-maintenance, and system-generated events. Examples of removed events include log report viewed, user profile viewed, role assigned, calendar event created, grade item updated, and user enrolled in course. After filtering, 369,592 learner-generated activity records from 760 users were retained for analysis. Each event was also mapped to its source course because some Moodle contexts such as quiz and assignment events are not stored directly as course contexts in the exported log.

Table 1. Dataset summary after preprocessing

Dataset element	Value
Raw Moodle logs	402,290
Filtered learner records	369,592
Course represented	20
Unique users	760
User-course instance	1,102
Observation	19 Feb 2025 - 2 Jun 2025

2.2. Feature construction

The cleaned logs were aggregated at the user-course level. This unit of analysis was selected because a learner may appear in more than one course, and engagement behavior is better interpreted within the context of a specific course. Eight behavioral indicators were constructed from the cleaned event logs as shown in Table 2.

Table 2. Behavioral features derived from Moodle logs

Feature	Operational definition
total_evemts	All learner events in course
active_days	Distinct access days
course_viewed	Course page access
module_viewed	Module/section/completion views
quiz_activity	Quiz attept/reviews/submission

assignment_activity	Assignment upload/status action
file_resource_activity	File/URL/page/folder/book access
distinct_event_types	Event-name diversity

2.3. Engagement scoring and clustering

Each feature was normalized using min-max scaling. An engagement score was then calculated as a weighted sum of normalized features, as shown in (1). The weights were defined to emphasize overall activity, persistence across active days, quiz interaction, assignment participation, and module access while still considering resource access and event diversity.

$$E_i = 100 \times \sum_{j=1}^{m} \left(w_j \times x_{ij}^{norm} \right), \sum_{j=1}^{m} 1 \tag{1}$$

The weight vector used in this study was 0.18 for total events, 0.18 for active days, 0.12 for course views, 0.14 for module views, 0.16 for quiz activity, 0.12 for assignment activity, 0.06 for file/resource activity, and 0.04 for event diversity. For clustering, log-transformed feature values were standardized and grouped using K-Means. Cluster validity was evaluated using silhouette score and Davies-Bouldin index. Although two clusters produced the highest silhouette value, three clusters were retained because they provide more actionable pedagogical categories: low, moderate, and high engagement.

3. RESULTS AND DISCUSSION

3.1. Moodle activity distribution

The cleaned dataset shows that Moodle activity was not evenly distributed across courses. The largest activity volume was found in Pengolahan Sinyal Digital A, followed by Pengolahan Sinyal Digital B and Elektronika dan Instrumentasi C. This pattern suggests that course design, assessment intensity, and the number of active learners may strongly influence the volume of LMS traces generated by a course.

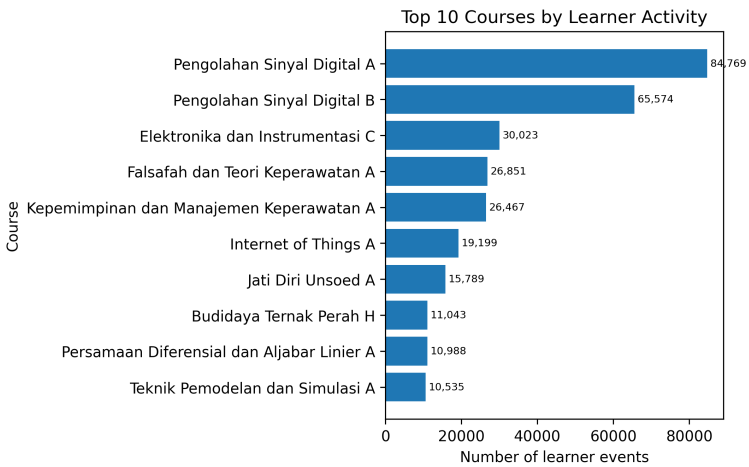

Figure 2. Top 10 courses by number of learner-generated Moodle events

Table 3. Dominant event names in the cleaned learner logs

Event name	Count	Share
Quiz attempt viewed	87,900	23.78%
Quiz attempt update	86,169	23.31%
Course module viewed	67,681	18.31%
Course viewed	42,584	11.52%
Submission status viewed	16,971	4.59%
Quiz attempt reviewed	6,878	1.86%
Quiz access was prevented	6,706	1.81%
Peerwork peer grade	6,570	1.78%

The dominant event types were quiz attempt viewed, quiz attempt updated, course module viewed, and course viewed. Quiz-related interactions formed the largest component of activity, indicating that assessment activity was a central driver of LMS interaction. Figure 3 confirms that quiz events were much more frequent than ordinary access events and assignment-related events.

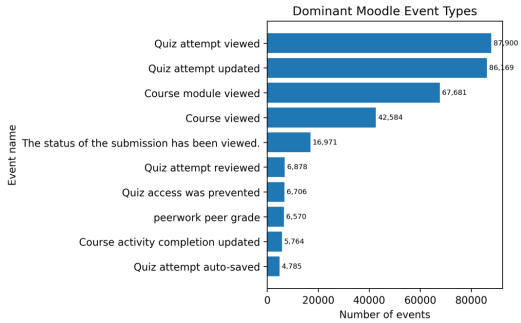

Figure 3. Dominant Moodle event types after filtering administrative records

3.2. User-course engagement characteristics

After aggregation, the analysis produced 1,102 user-course engagement instances. The average user-course record contained 328.42 events, 13.55 active days, 38.64 course views, 70.16 module-related views, 199.87 quiz-related activities, and 54.03 assignment-related activities. The median values were lower than the means for most features, indicating that activity was positively skewed: a smaller number of users generated very large volumes of interaction, while many users interacted at moderate or low frequency.

Table 4. Descriptive statistics of engagement features

Feature	Mean	Median	Max
total_events	328.42	172.00	2,118.00
active_days	13.55	12.00	53.00
course_viewed	36.64	28.00	252.00
module_viewed	70.16	51.00	384.00
quiz_activity	199.87	41.00	1,731.00
assignment_activity	54.03	32.00	470.00
file_resource_activity	17.24	14.00	109.00
distinct_event_types	14.15	15.00	28.00

3.3. Engagement clustering

Cluster validation was examined for two to five clusters. The two-cluster solution achieved the highest silhouette score, but the three-cluster solution was selected because it offers a more practical interpretation for academic monitoring. In intervention-oriented learning analytics, three categories are often easier to translate into institutional action: low engagement for immediate attention, moderate engagement for monitoring, and high engagement for maintenance and enrichment.

Table 5. Cluster validity metrics

k	Sil.	DBI	Cluster size
2	0.368	1.062	387 / 715
3	0.306	1.087	318 / 198 / 586
4	0.290	1.157	203 / 419 / 151 / 249
5	0.315	1.111	294 / 401 / 220 / 119 / 68

Table 6. Mean characteristics of the three engagement profiles

Profile	n	Events	Days	Quiz	Assign.	Score
Low engagement	198	82.88	3.87	61.66	4.71	4.38
Moderate engagement	586	164.26	11.87	63.29	46.24	12.71
High engagement	318	785.80	22.07	537.60	99.11	55.03

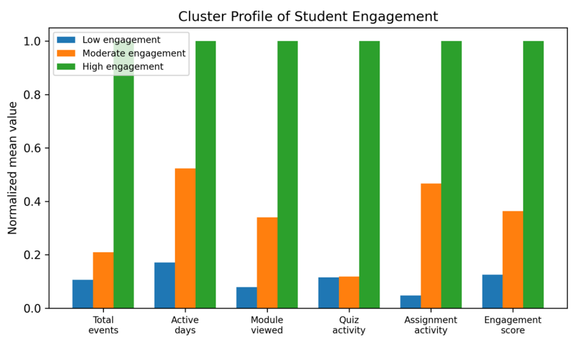

Figure 4. Normalized profile of low, moderate, and high engagement clusters

3.4. Discussion

The results demonstrate that raw Moodle logs can be transformed into interpretable engagement indicators through a relatively simple learning analytics pipeline. The dominance of quiz activity suggests that students interacted most intensively with assessment-related functions. This finding is consistent with previous learning analytics studies showing that assessment interactions, temporal engagement patterns, and LMS-based behavioral features are important indicators of online learning behavior [1], [6], [7], [12].

The high engagement cluster generated substantially more events, active days, module views, quiz actions, and assignment interactions than the other groups. This profile indicates sustained and diverse participation rather than isolated login behavior. In contrast, the low engagement cluster showed very limited active days and lower module and assignment interaction. This pattern can be useful for lecturers because it points to learners who may require reminders, academic advising, or additional support before the end of the course.

The moderate cluster was the largest group, representing learners who interacted with the LMS but did not reach the intensity of the high engagement cluster. For teaching practice, this group may be important because small instructional interventions, such as timely feedback, course reminders, and clearer assessment navigation, may increase their engagement. Learning analytics dashboards and LMS-based interventions have been proposed in previous studies as practical mechanisms for supporting such actions [8], [11], [14].
A key limitation of this study is that the dataset did not include final grades, demographic data, or lecturer-validated risk labels. Therefore, the findings should be interpreted as behavioral engagement profiles, not as proof of academic success or failure. Future work should integrate gradebook data, attendance records, and course completion data to evaluate whether the engagement profiles are associated with learning outcomes. Ethical data governance is also necessary because LMS logs can contain personally identifiable information such as names and IP addresses.

3.5. Temporal pattern of LMS access

Temporal analysis was conducted to complement the aggregate engagement profiles. Aggregate counts are useful for describing the volume of interaction, but they do not show whether learners accessed the LMS consistently, only during assessment periods, or irregularly across the semester. Therefore, temporal information is important for interpreting Moodle event logs and for designing timely academic interventions [1], [7], [13].

Figure 5 shows that activity was not constant across the observation period. Several weeks produced considerably higher interaction volumes, which is consistent with the presence of quiz attempts, assignment submissions, and course module views. This pattern suggests that LMS engagement was strongly shaped by the assessment calendar and course design. Therefore, lecturers should not interpret a sudden increase in activity as sustained engagement without checking whether the increase was caused by a specific quiz or assignment deadline.

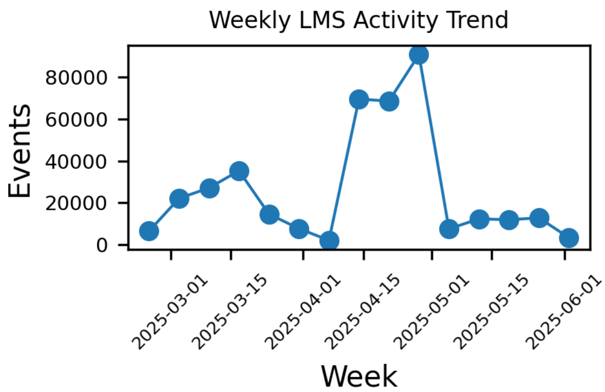

Figure 5. Weekly trend of learner-generated LMS activity

The hourly distribution in Figure 6 provides an additional perspective on learning behavior. A high number of activities outside formal class hours may indicate that students use the LMS as an asynchronous learning space rather than only as a classroom-support tool. This finding is important for academic monitoring because it shows that learning support should not be limited to synchronous meeting times. Technical accessibility, clarity of instructions, and availability of materials become essential when students access the LMS at different times of the day.

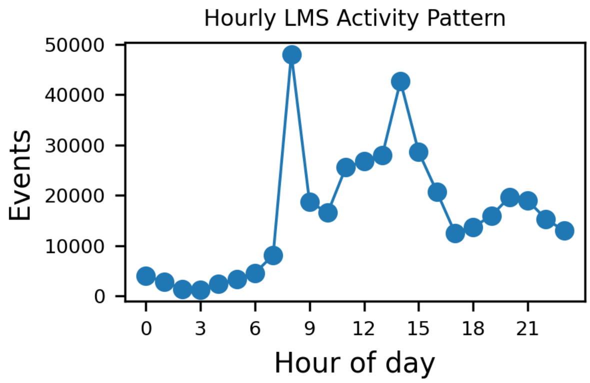

Figure 6. Hourly pattern of Moodle learner activity

3.6. Activity category contribution

The feature construction stage grouped the Moodle events into categories that reflect different dimensions of learning behavior. Course views represent general access to the course page, module views indicate navigation across learning activities, quiz activity reflects assessment interaction, assignment activity represents task-related participation, and file-resource activity indicates access to learning materials. This categorization helps transform technical event names into pedagogically meaningful indicators.

Table 7 and Figure 7 show that quiz activity was the largest contributor to the aggregated feature set. The high contribution of quiz activity means that many student actions were generated during assessment interaction, especially when attempts were viewed, updated, reviewed, or submitted. Module views and course views also contributed substantially, showing that learners repeatedly navigated course pages and learning activities. Assignment activity appeared as another important component because it captures submission-status checks, upload actions, and task-related viewing behavior.

Table 7. Aggregated contribution of Moodle activity categories

Activity category	Count	Share
quiz activity	220,252	52.58%
module viewed	77,321	18.46%
assignment activity	59,544	14.21%
course viewed	42,584	10.17%
file resource activity	19,001	4.54%
forum activity	193	0.05%
attendance activity	0	0.00%

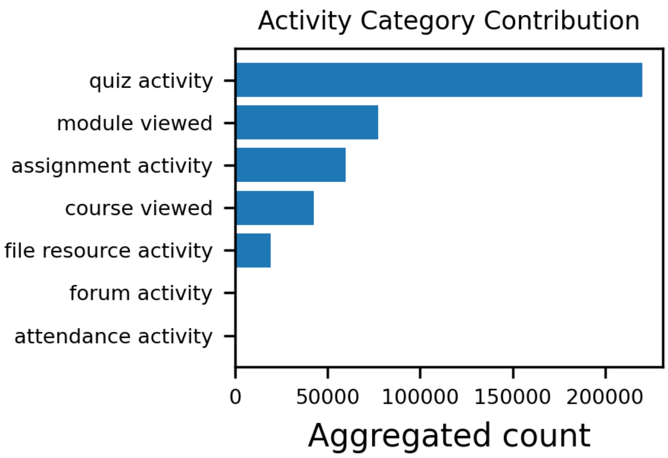

Figure 7. Contribution of Moodle activity categories after aggregation

These results show that engagement profiling should not depend only on the number of logins or course views. A learner can open the course page several times without engaging with assessment or content, while another learner may generate fewer events but interact meaningfully with quizzes, assignments, or learning materials. Previous reviews also emphasize that LMS traces should be interpreted carefully because behavioral indicators represent observable engagement rather than the full quality of learning [2], [5].

3.7. Course-level engagement variation

Because the dataset was collected from multiple courses, engagement profiles were also examined at course level. This analysis is necessary because different courses may require different levels of LMS interaction. A course with frequent quizzes and online activities naturally generates more event records than a course that uses Moodle mainly for uploading files. Therefore, course-level interpretation prevents unfair comparison between courses with different instructional designs.

Table 8 presents the distribution of low, moderate, and high engagement profiles in the ten largest courses by user-course instance. The table shows that the proportion of low and high engagement was not identical across courses. This variation may be influenced by course structure, number of assessments, lecturer use of Moodle, student cohort characteristics, and the extent to which learning activities were conducted online. Figure 8 visualizes the share of low and high engagement groups across selected courses.

Table 8. Distribution of engagement profiles across selected courses

Course	Low	Mod.	High	Low %	High %
*** A	10	45	0	18.1	0.0
Industri Persusuan K	57	21	1	69.1	3.4
Teknik Pemodelan dan Simulasi	2	47	15	3.1	23.4
Elektronika dan Instrumentasi	2	22	26	3.8	48.6
Sistem Pertanian Cerdas B	15	29	2	28.3	5.1
Persamaan Diferensial dan Aljabar	4	50	3	6.5	32.8
Teknik Pemodelan dan Simulasi	2	42	18	3.2	29.6
Sistem Pertanian Cerdas A	6	52	3	9.8	4.9
Falsafah dan Teori Keperawatan	12	6	48	36.7	69.0
Pengolahan Sinyal Digital A	1	0	57	1.7	98.3

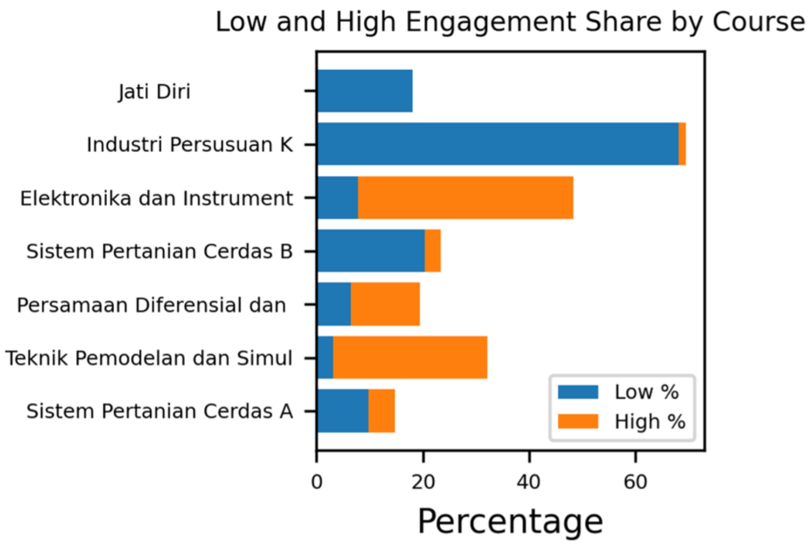

Figure 8. Low and high engagement share across selected courses

The course-level results have practical implications for quality assurance. If a course has many low-engagement students, lecturers and program managers should first examine whether the Moodle course is actively used, whether learning instructions are clear, and whether activities are scheduled consistently. A low engagement profile does not automatically indicate low student motivation; it may also reflect limited use of Moodle in that course. For this reason, learning analytics should be interpreted together with pedagogical context.

3.8. Course activity intensity

In addition to cluster distribution, average activity intensity was calculated for each course. This analysis focuses on how many events, active days, quiz actions, and assignment actions were generated per user-course record. Table 9 shows that some courses had substantially higher event intensity than others. Courses with higher average events and active days may have used Moodle more extensively for learning activities, assessment, or continuous interaction.

The difference in course intensity also explains why engagement scoring should be used carefully. If a student belongs to a course with many online quizzes, the expected number of quiz-related events will be higher than in a course that relies on offline assessment. Therefore, future versions of the model can apply course-normalized scoring, where each learner is compared primarily with peers in the same course. This approach would reduce the risk of overestimating or underestimating engagement simply because of different course designs.

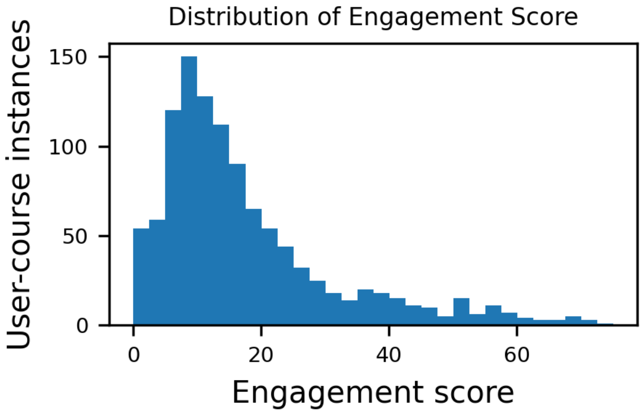

Figure 9. Distribution of engagement scores at user-course level

The engagement score distribution in Figure 9 confirms that the dataset was not normally distributed. Most user-course instances were concentrated in the lower and middle score ranges, while a smaller group of highly active learners produced much higher scores. This skewed pattern supports the use of clustering because clustering can separate behavioral groups without forcing a fixed linear threshold. However, the score remains useful for ranking and monitoring because it provides a continuous indicator before categorical grouping is applied.

3.9. Research contribution

The contribution of this study is positioned at the intersection of learning analytics practice and educational data mining methodology. The dataset demonstrates that raw LMS exports can be converted into meaningful engagement profiles through a clear sequence of preprocessing, aggregation, scoring, clustering, and interpretation. The study does not propose a new machine learning algorithm, but it provides an applied analytical model that can be directly replicated in institutions using Moodle.

Table 9 summarizes the contribution of the study. The contribution is not limited to technical clustering. It includes data-level contribution through the use of raw multi-course logs, methodological contribution through an integrated workflow, interpretive contribution by distinguishing engagement profiling from achievement prediction, practical contribution through lecturer-facing actions, and governance contribution through privacy-aware reporting.

Table 9. Summary of research contributions

Area	Contribution
Data	Raw multi-course Moodle logs
Method	Filtering, features, score, clusters
Interpretation	Engagement profile, not grade claim
Practice	Lecturer-facing monitoring actions
Governance	Anonymized aggregate reporting

This contribution is relevant for universities that already possess LMS log data but have not converted them into academic decision-support information. The results show that even without gradebook integration, Moodle logs can support descriptive analytics, engagement profiling, dashboard design, and institutional monitoring. With additional academic outcomes, the same workflow can be expanded toward early warning systems and adaptive learning support [11], [15], [17].

4. CONCLUSION

This study developed a Moodle log-based learning analytics workflow for profiling student engagement across 20 courses. From 402,290 raw log records, 369,592 learner-generated activities were extracted and transformed into behavioral features at the user-course level. The analysis shows that quiz activities, course/module views, and assignment-related events are the dominant indicators of LMS interaction. K-Means clustering produced three interpretable profiles: low, moderate, and high engagement. Additional temporal and course-level analyses show that engagement patterns vary across weeks, hours, activity categories, and course designs. The proposed workflow can support lecturers and academic managers in monitoring participation, designing dashboard indicators, and identifying learners who may require further support. The study also emphasizes that engagement profiling should not be overinterpreted as academic performance prediction when gradebook data are unavailable. Future studies should integrate final grades, attendance, completion status, and lecturer-validated intervention labels to evaluate whether early LMS engagement indicators can predict academic outcomes and improve the effectiveness of student-support actions.

ACKNOWLEDGMENTS

This section should acknowledge individuals who provided personal assistance to the work but do not meet the criteria for authorship, detailing their contributions. It is imperative to obtain consent from all individuals listed in the acknowledgments.

REFERENCES

[1] D. Rotelli and A. Monreale, "Processing and understanding Moodle log data and their temporal dimension," Journal of Learning Analytics, vol. 10, no. 2, pp. 126-141, 2023, doi: 10.18608/jla.2023.7867.
[2] N. Bergdahl, M. Bond, J. Sjöberg, M. Dougherty, and E. Oxley, "Unpacking student engagement in higher education learning analytics: A systematic review," International Journal of Educational Technology in Higher Education, vol. 21, art. 63, 2024, doi: 10.1186/s41239-024-00493-y.
[3] N. A. Johar, S. N. Kew, Z. Tasir, and E. Koh, "Learning analytics on student engagement to enhance students' learning performance: A systematic review," Sustainability, vol. 15, no. 10, art. 7849, 2023, doi: 10.3390/su15107849.
[4] S. Fan, L. Chen, M. Nair, S. Garg, S. Yeom, G. Kregor, Y. Yang, and Y. Wang, "Revealing impact factors on student engagement: Learning analytics adoption in online and blended courses in higher education," Education Sciences, vol. 11, no. 10, art. 608, 2021, doi: 10.3390/educsci11100608.
[5] Q. Wang and A. Mousavi, "Which log variables significantly predict academic achievement? A systematic review and meta-analysis," British Journal of Educational Technology, vol. 54, no. 1, pp. 142-191, 2023, doi: 10.1111/bjet.13282.
[6] C. Lu and M. Cutumisu, "Online engagement and performance on formative assessments mediate the relationship between attendance and course performance," International Journal of Educational Technology in Higher Education, vol. 19, art. 2, 2022, doi: 10.1186/s41239-021-00307-5.
[7] R. M. Santos and R. Henriques, "Accurate, timely, and portable: Course-agnostic early prediction of student performance from LMS logs," Computers and Education: Artificial Intelligence, vol. 5, art. 100175, 2023, doi: 10.1016/j.caeai.2023.100175.
[8] G. Ramaswami, T. Susnjak, and A. Mathrani, "Effectiveness of a learning analytics dashboard for increasing student engagement levels," Journal of Learning Analytics, vol. 10, no. 3, pp. 115-134, 2023, doi: 10.18608/jla.2023.7935.
[9] S. Kim, S. Cho, J. Y. Kim, and D.-J. Kim, "Statistical assessment on student engagement in asynchronous online learning using the k-means clustering algorithm," Sustainability, vol. 15, no. 3, art. 2049, 2023, doi: 10.3390/su15032049.
[10] D. Amo, S. Cea, N. M. Jimenez, P. Gómez, and D. Fonseca, "A privacy-oriented local web learning analytics JavaScript library with a configurable schema to analyze any edtech log: Moodle's case study," Sustainability, vol. 13, no. 9, art. 5085, 2021, doi: 10.3390/su13095085.
[11] Z. Pan, L. Biegley, A. Taylor, and H. Zheng, "A systematic review of learning analytics: Incorporated instructional interventions on learning management systems," Journal of Learning Analytics, vol. 11, no. 2, pp. 52-72, 2024, doi: 10.18608/jla.2023.8093.
[12] M. Fazil, A. Rísquez, and C. Halpin, "A novel deep learning model for student performance prediction using engagement data," Journal of Learning Analytics, vol. 11, no. 2, pp. 23-41, 2024, doi: 10.18608/jla.2024.7985.
[13] M. R. Marcolino, T. R. Porto, T. T. Primo, R. Targino, V. Ramos, E. M. Queiroga, R. Munoz, and C. Cechinel, "Student dropout prediction through machine learning optimization: Insights from Moodle log data," Scientific Reports, vol. 15, art. 9840, 2025, doi: 10.1038/s41598-025-93918-1.
[14] L. Chen, X. Geng, M. Lu, A. Shimada, and M. Yamada, "How students use learning analytics dashboards in higher education: A learning performance perspective," SAGE Open, vol. 13, no. 3, 2023, doi: 10.1177/21582440231192151.
[15] H. M. M. Ahmed, H. A. El-Sabagh, and D. M. Elbourhamy, "Effect of gamified, mobile, cloud-based learning management system on student engagement and achievement," International Journal of Educational Technology in Higher Education, vol. 22, art. 49, 2025, doi: 10.1186/s41239-025-00541-1.
[16] M. Khalil, P. Prinsloo, and S. Slade, "The use and application of learning theory in learning analytics: A scoping review," Journal of Computing in Higher Education, vol. 35, pp. 573-594, 2023, doi: 10.1007/s12528-022-09340-3.
[17] Y. Soepriyanto, R. P. Nugroho, M. H. A. Nahri, D. W. Kesuma, and M. Setiasih, "From logs to insights: A comprehensive framework for data-driven learning insights," Jurnal Inovasi Teknologi Pendidikan, vol. 12, no. 1, pp. 40-49, 2025, doi: 10.21831/jitp.v12i1.77432.