Educational Data Mining of VLE Interaction Logs for Explaining Student Assessment Outcomes: A K-Means Clustering Study

Yasanthika Mathotaarachchi
Department of Software Engineering & Computer Security
NSBM Green University
Homagama, Sri Lanka
yasanthika.m@nsbm.ac.lk

***Abstract*—Virtual Learning Environments (VLEs) generate extensive traces of learner interaction; however, educators and institutions often lack interpretable analytical approaches to determine which online learning activities meaningfully contribute to assessment success. This study applies Educational Data Mining (EDM) techniques to investigate the relationship between students' online learning behaviors and assessment outcomes using large-scale VLE interaction data. The Open University Learning Analytics Dataset (OULAD), containing detailed student, assessment, and activity records across multiple course presentations, is used as the empirical basis of the analysis. A consolidated learner–activity–assessment dataset is constructed through relational integration and preprocessing of multiple OULAD tables. K-Means clustering is applied to identify distinct engagement–performance profiles without relying on predefined outcome labels. The analysis results in four clusters with heterogeneous engagement characteristics, among which one cluster emerges as dominant, capturing the highest proportion of meaningful learning interactions and demonstrating a strong association with higher assessment performance. Further analysis examines activity-type distributions across clusters to evaluate the relative influence of different online learning activities. The results indicate that engagement with structured learning resources and core content activities exhibits the strongest relationship with academic success, while activities supporting formative practice and interaction contribute positively when combined with content-focused engagement. In contrast, excessive interaction with navigational or supplementary activities shows limited association with improved performance. These findings suggest that engagement quality and activity semantics are more informative indicators of learning effectiveness than overall activity volume. The study provides interpretable, cluster-based evidence to support data-driven instructional decision-making and highlights the importance of guiding students toward high-impact learning activities in higher education contexts.**

Keywords— Educational Data Mining, Learning Analytics, LMS logs, K-Means clustering, student performance

I. INTRODUCTION

The widespread adoption of Learning Management Systems (LMS) and Virtual Learning Environments (VLEs) has led to continuous streams of educational data capturing student interactions with learning resources, forums, quizzes, and assessment workflows. Learning Analytics (LA) and Educational Data Mining (EDM) have emerged as key approaches for transforming such data into actionable insights that support instructional design and learner success [1], [2]. A central challenge, however, is moving beyond descriptive analytics toward evidence-based identification of which VLE activities most strongly influence assessment outcomes, particularly in large-scale learning environments where manual monitoring is infeasible.

Prior work has shown that VLE logs can capture meaningful proxies of participation and learning community behavior [3], and systematic reviews highlight LA/EDM's growing role in predicting and understanding student performance and engagement patterns in higher education [1], [2]. Nonetheless, institutions frequently face two practical gaps:

1. translating clickstream and activity logs into **interpretable learner profiles**, and
2. providing educators with **actionable guidance** on which activity categories to encourage.

This study addresses these gaps using clustering-based EDM. Clustering is suitable when labels are incomplete or when the goal is exploratory discovery of engagement-performance groups. K-Means remains popular due to its simplicity and interpretability in educational contexts [3]. Additionally, OULAD has become a widely used benchmark dataset for LA/EDM studies, enabling reproducibility and comparative evaluation [4], [5].

This paper makes three contributions. First, it applies large-scale Educational Data Mining to OULAD to model relationships between VLE activity engagement and student assessment outcomes [5]. Second, it provides an interpretable K-Means clustering analysis that identifies four engagement–performance profiles and ranks activity types into most important, important, and average categories using cluster-weighted evidence. Third, it offers practical insights for guiding students toward high-impact learning activities in higher education.

The objective of this study is to investigate how different categories of VLE learning activities relate to student assessment outcomes using clustering-based Educational Data Mining. Specifically, the study aims to: 1) identify interpretable engagement–performance profiles from OULAD using K-Means clustering; 2) determine which VLE activity types are most strongly associated with higher assessment performance; and 3) provide an interpretable basis for data-informed instructional decision-making in higher education.

The contribution of this study lies not in proposing a new clustering algorithm or a new benchmark dataset, but in using clustering as an explanatory tool to derive interpretable, activity-level insights that are pedagogically meaningful.

II. Related Work

Educational Data Mining (EDM) and Learning Analytics (LA) have become prominent research domains aimed at extracting meaningful insights from educational data to improve learning outcomes, instructional strategies, and institutional decision-making. With the widespread adoption of Virtual Learning Environments (VLEs) and Learning Management Systems (LMSs), large volumes of learner interaction data are continuously generated, enabling data-driven analysis of student engagement and performance [1], [2].

A. Educational Data Mining for Student Performance Analysis

Early research in EDM primarily focused on predicting student performance using supervised learning techniques such as decision trees, Bayesian classifiers, neural networks, and ensemble models. Romero and Ventura demonstrated that data mining techniques could uncover pedagogically meaningful patterns from LMS data, supporting instructors in understanding learner behavior and refining instructional strategies [3]. Similarly, Mehboob employed classification-based approaches to predict student failure risk and identify contributing factors using academic history data [3].

While these predictive approaches achieve reasonable accuracy, they are largely outcome-oriented and often function as black-box models. As highlighted in recent systematic reviews, many prediction-focused studies emphasize accuracy over interpretability, limiting their usefulness for pedagogical decision-making and instructional design [1], [4].

A key knowledge gap is that, despite extensive work on performance prediction; there is limited empirical research explaining how and which specific online learning activities influence assessment outcomes. Most studies identify who is at risk, but not why particular engagement behaviors lead to success or failure.

B. Clustering-Based Learner Profiling in Learning Analytics.

Unsupervised learning techniques, particularly clustering, have been widely applied to profile learner behavior and engagement patterns without relying on labeled outcomes. Jamesmanoharan et al. applied K-Means clustering to group students based on academic performance, demonstrating its effectiveness in identifying distinct performance-based learner groups [5]. López et al. further explored clustering-based classification approaches using Moodle forum activity data to differentiate between pass and fail outcomes [6].

Although clustering has proven useful for learner segmentation, most existing studies use clustering primarily as a descriptive grouping mechanism rather than an explanatory tool for evaluating the relative contribution of different learning activities. Cluster interpretation is often limited to performance averages or participation levels, without systematically linking clusters to activity-level engagement evidence. In contrast, the present study uses clustering not only to identify engagement–performance profiles, but also to derive an interpretable ranking of VLE activity types based on cluster-weighted evidence. This shifts the emphasis from learner grouping alone toward explanatory and pedagogically meaningful analytics.

A key methodological gap is the lack of clustering-based studies that leverage cluster characteristics to quantitatively evaluate and rank online learning activities based on their contribution to assessment performance. Most clustering studies stop learning grouping and do not translate results into actionable pedagogical insights.

C. Analysis of VLE Activity Logs and Engagement Metrics

Several studies have examined LMS and VLE activity logs to understand student engagement and online learning behavior. Black et al. demonstrated that LMS log data could be used to analyze online learning communities and participation patterns [6]. Guo et al. analyzed learner interaction with video-based content and found that engagement quality, rather than duration alone, significantly affects learning outcomes [7].

Recent learning analytics literature consistently reports that high activity volume does not necessarily correspond to improved academic performance. Aggregate measures such as total clicks or time spent are insufficient proxies for learning, as they fail to capture the pedagogical value of different activity types [2], [7].

Another knowledge gap is that, although prior studies acknowledge the importance of engagement quality, few provide a data-driven categorization of VLE activity types based on their empirical relationship with assessment outcomes. Existing work often relies on correlation analysis or descriptive statistics without systematically distinguishing high-impact learning activities from low-impact interactions.

D. Technological Gaps in Educational Data Mining Tools

Despite advances in EDM methodologies, their integration into practical educational tools remains limited. Merceron and Yacef proposed data mining tools to assist instructors in identifying learning patterns and pedagogical issues; however, such tools often require technical expertise and are not easily accessible to educators without data mining backgrounds [8]. Modern LMS platforms offer built-in analytics dashboards, but these dashboards typically provide surface-level statistics and lack deeper analytical intelligence [1], [8].

Identified Gap is There is a notable absence of lightweight, educator-friendly systems that translate EDM results into **actionable visual analytics**. Most existing tools either remain research prototypes or present analytics in forms that are difficult for educators to interpret and act upon in real teaching contexts.

E. Synthesis of Research Gaps and Study Positioning

Based on the reviewed literature, four critical gaps are identified:

Knowledge Gap: Insufficient understanding of how different categories of online learning activities influence assessment outcomes beyond general engagement metrics.

Methodological Gap: Limited use of clustering results to derive interpretable, activity-level insights that explain variations in student performance.

Technological Gap: Lack of practical, web-based tools that operationalize EDM findings for educators and students.

Practical Gap: Absence of decision-support mechanisms that guide learners toward high-impact activities rather than encouraging indiscriminate platform usage.

To address these gaps, this study applies K-Means clustering to large-scale VLE interaction data from the Open University Learning Analytics Dataset (OULAD) [9], [10], systematically analyzes cluster-weighted activity engagement to rank learning activities based on their impact on assessment outcomes.

III. MATERIALS AND METHODS

A. Dataset

This study utilizes the Open University Learning Analytics Dataset (OULAD), which contains anonymized student demographic data, assessment outcomes, and detailed Virtual Learning Environment (VLE) interaction logs collected across multiple modules and course presentations [4], [5]. OULAD was originally donated to the UCI Machine Learning Repository and later curated and published in Scientific Data, making it a widely accepted benchmark dataset for Educational Data Mining and Learning Analytics research [4], [5].

The dataset comprises seven relational tables, including student information, student registration, student assessment, student VLE interactions, assessments, courses, and VLE activity definitions. These tables collectively capture both academic performance indicators and fine-grained online learning behavior across different learning contexts.

B. Feature Construction and Integrated Table

Following a relational integration approach, a consolidated dataset was constructed by merging multiple OULAD tables using structured query operations, consistent with the data modeling pipeline as described in [8]. The integration process resolves schema mismatches and aligns student identifiers, module codes, and assessment references across tables.

The final integrated table includes key attributes such as:

module and presentation identifiers, student identifiers, VLE activity type, aggregated engagement measure (sum of clicks), assessment identifiers, assessment scores, and final assessment result (Pass/Fail/Withdrawn).

This consolidated structure enables systematic cluster analysis by explicitly linking student engagement patterns with assessment performance, thereby supporting activity-level interpretation of learning outcomes [8].

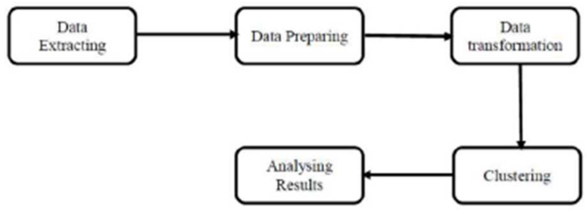

Fig.1. Data mining workflow including data integration, preprocessing, clustering, and analysis.

C. Data Preprocessing

Data preprocessing was conducted to ensure data quality, consistency, and suitability for clustering analysis. First, the seven OULAD tables were merged using relational joins in MySQL to create a unified analytical dataset. Second, incomplete, inconsistent, and noisy records were examined and removed where necessary to improve data quality. Third, missing values were handled using WEKA preprocessing filters, and non-informative attributes were excluded. Finally, engagement-related variables were transformed into a consistent analytic structure prior to clustering to ensure meaningful comparison of feature values during distance-based clustering. These preprocessing steps reduced noise and improved the stability and interpretability of the clustering results. All preprocessing procedures were implemented using WEKA for data preparation and validation, supported by MySQL for relational data handling.

D. Clustering Model: K-Means

K-Means clustering was employed to identify distinct engagement–performance profiles among students. The algorithm partitions the dataset into K clusters by minimizing the within-cluster sum of squared Euclidean distances. Given feature vectors x_i, K-Means iteratively assigns each data point to the nearest centroid and updates centroid positions until convergence.

K-Means was selected due to its computational efficiency, interpretability, and widespread use in EDM studies focusing on learner behavior profiling [4]. The objective of clustering in this study is not prediction, but exploration discovery of engagement patterns and their relationship with assessment outcomes.

$$J = \sum_{i=1}^{m} \sum_{k=1}^{K} w_{ik} \| x^i - \mu_k \|^2$$

Fig. 2. Overview of the K-Means clustering process applied to VLE interaction data.

Several values of K were explored during experimentation, and K = 4 was selected as the final solution because it provided the most interpretable separation of engagement–performance patterns without producing overly coarse or overly fragmented cluster structures. As the analysis is exploratory, the choice of K was guided primarily by the interpretability of cluster profiles and the meaningful differentiation of activity distributions across clusters, consistent with the explanatory objective of the study [8]. The resulting cluster distribution is as follows:

TABLE I. DISTRIBUTION OF STUDENT INSTANCES ACROSS CLUSTERS

Cluster ID	Number of Instances	Percentage (%)
Cluster 0	295,011	28
Cluster 1	148,941	14
Cluster 2	350,298	33
Cluster 3	254,325	24

Among the four clusters, Cluster 2 contains the highest proportion of records and shows the strongest association with meaningful learning engagement and assessment success. Consequently, this cluster was selected as the primary focus for deeper activity-effect analysis [8]. The analysis retained features that directly connect learner interaction behavior with assessment outcomes, including module and presentation identifiers, student identifiers, VLE activity types, aggregated engagement measures such as sum of clicks, assessment identifiers, assessment scores, and final assessment results. These attributes were selected because they jointly capture behavioral engagement and academic performance while preserving interpretability for cluster-level analysis.

E. *Activity Importance Categorization*

To identify which online learning activities contribute most significantly to academic performance, activity-type distributions were examined across all clusters, with particular emphasis on Cluster 2. Activity importance was derived using cluster-weighted engagement evidence, where activities with higher representation in the high-performing cluster were considered more influential.

Based on this analysis, VLE activities were categorized into three levels:

- Most Important: resource, subpage, oucontent, url
- Important: forumng, quiz, page, questionnaire
- Average: oucollaborate, homepage, glossary, dataplus, ouelluminate, sharedsubpage, externalquiz, ouwiki, dualpane, repeatactivity, older, htmlactivity

This categorization reflects the empirical relationship between activity engagement and assessment outcomes rather than raw usage frequency. The results indicate that purposeful engagement with content-centric and assessment-aligned activities has a stronger impact on performance than indiscriminate interaction with the VLE [8].

IV. RESULTS

A. *Cluster Profiles and Key Findings*

The application of the K-Means clustering algorithm to the integrated OULAD dataset resulted in four clusters with uneven instance distributions. Among these, Cluster 2 captured the largest proportion of records, accounting for approximately 33% of the total dataset, while the remaining instances were distributed across Cluster 0 (28%), Cluster 1 (14%), and Cluster 3 (24%). The imbalance in cluster sizes indicates heterogeneity in student engagement patterns within the Virtual Learning Environment (VLE).

Detailed inspection of cluster characteristics reveals that Cluster 2 represents students who demonstrate the highest level of meaningful engagement with online learning activities, particularly those aligned with instructional content and assessment preparation. Students in this cluster exhibit sustained interaction with learning resources, structured content, and assessment-related activities, which collectively correspond to stronger academic performance outcomes. Consequently, Cluster 2 is treated as the primary cluster of interest for subsequent activity-effect analysis.

A key insight emerging from the clustering results is that high overall interaction volume alone does not guarantee higher assessment performance. Although some clusters exhibit substantial click activity, these interactions are often concentrated on low-impact or navigational activities. In contrast, students in Cluster 2 engage more purposefully with pedagogically meaningful activities. This finding highlights that the quality and relevance of engagement outweigh raw activity volume, reinforcing the need to interpret VLE engagement through activity semantics rather than aggregate usage metrics alone. This observation aligns with recent Learning Analytics research emphasizing that engagement quality is a stronger indicator of learning effectiveness than mere time-on-task or click counts [2].

```
Class attribute: activitytype
Classes to Clusters:
      0      1      2      3   <-- assigned to cluster
 295011 148941 346594 254325 | resource
      0      0    996      0 | oucontent
      0      0    886      0 | url
      0      0     22      0 | homepage
      0      0   1055      0 | subpage
      0      0     21      0 | glossary
      0      0    194      0 | forumng
      0      0     82      0 | oucollaborate
      0      0     28      0 | dataplus
      0      0    127      0 | quiz
      0      0     21      0 | ouelluminate
      0      0      3      0 | sharedsubpage
      0      0     61      0 | questionnaire
      0      0    102      0 | page
      0      0     26      0 | externalquiz
      0      0     49      0 | ouwiki
      0      0     20      0 | dualpane
      0      0      5      0 | repeatactivity
      0      0      2      0 | folder
      0      0      4      0 | htmlactivity
```

Fig. 3. cluster relationship with online activities

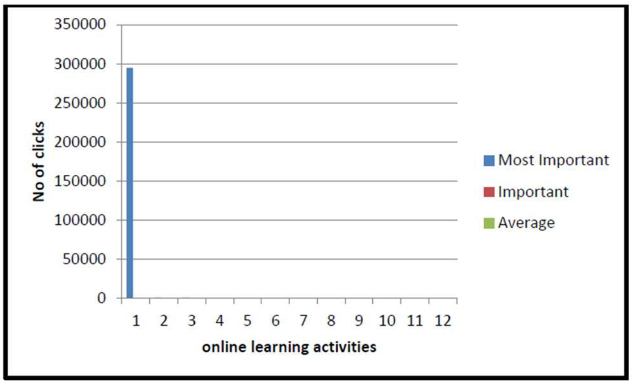

Fig. 4. Distribution of online learning activities within the high-performing cluster.

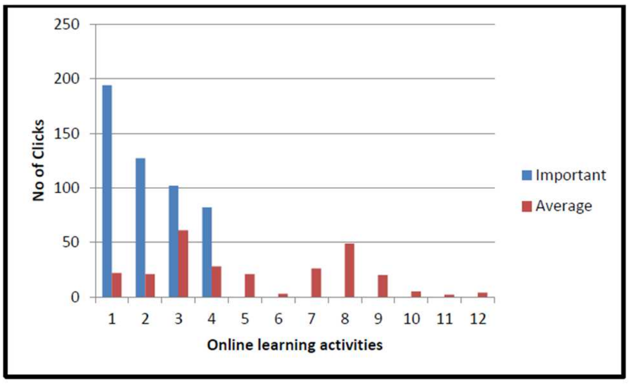

Fig. 5. Comparison of important and average activity categories across clusters.

B. *Activity Type Importance*

To further understand how different online learning activities influence assessment outcomes, activity-type distributions were analyzed across all clusters, with particular emphasis on Cluster 2. Based on cluster-weighted engagement evidence, VLE activities were systematically categorized into Most Important, Important, and Average groups.

Activities categorized as Most Important, including resource, subpage, oucontent, and url exhibit the highest concentration of interactions within Cluster 2. Among these, resources emerge as the most frequently engaged activity type, indicating that direct interaction with learning materials plays a critical role in academic success. These activities are primarily content-centric and support conceptual understanding, revision, and independent learning, which directly contribute to assessment preparedness.

The Important category includes activities such as quiz, forumng, page, and questionnaire. These activities facilitate formative assessment, peer interaction, and reflective learning. Their strong representation within the high-performing cluster suggests that assessment-aligned practice and interactive learning opportunities reinforce conceptual understanding and support improved performance. This finding is consistent with prior studies that highlight the importance of formative assessment and collaborative learning in online education [1], [2].

Activities classified as Average, including homepage, glossary, dataplus, ouelluminate, ouwiki, and similar tools are more evenly distributed across clusters and show a weaker association with assessment success. While these activities support navigation, communication, and supplementary learning, excessive reliance on them without sufficient engagement in high-impact activities does not appear to significantly enhance academic outcomes. This reinforces the conclusion that not all VLE activities contribute equally to learning effectiveness.

V. Discussion

This study provides interpretable, cluster-based empirical evidence demonstrating that different categories of Virtual Learning Environment (VLE) activities exhibit distinct relationships with assessment outcomes. While raw interaction counts (e.g., number of clicks) are often used as proxies for engagement, the results clearly indicate that activity semantics, that is, what students engage with, are more influential than how frequently they interact. Specifically, activities associated with resource review, structured content consumption, and purposeful navigation (e.g., resource, oucontent, url, and subpage) were consistently overrepresented within the high-performing cluster and were therefore ranked as most important activities. This suggests that engagement directed toward core instructional content plays a central role in supporting conceptual understanding and assessment readiness, as evidenced by the clustering outcomes.

The originality of this study lies in its explanatory use of clustering. Rather than employing K-Means solely for learner segmentation, the analysis uses cluster patterns to identify which categories of VLE activities are more strongly associated with higher assessment outcomes.

In contrast, activities that support practice, reflection, and interaction, such as quiz, forumng, and questionnaire, were categorized as important rather than dominant. This distinction is pedagogically meaningful: while these activities may not independently drive performance, they appear to reinforce learning when combined with strong content engagement. The findings thus support a multifaceted engagement model, where effective learning emerges from a balanced combination of content access and active learning behaviors. This interpretation is consistent with Learning Analytics and EDM literature that emphasizes the complementary roles of content interaction, formative assessment, and collaborative learning in online education [1], [2].

From an institutional and instructional design perspective, these results challenge the common assumption that increasing overall LMS usage will automatically lead to better academic outcomes. Instead, the evidence supports pathway-oriented interventions, where students are guided toward high-impact sequences of activities rather than encouraged to simply “spend more time” on the platform. This insight is particularly valuable for large-scale courses, where instructors may struggle to identify effective engagement strategies using traditional LMS analytics dashboards. By leveraging OULAD as an open benchmark dataset, this study also enables replication and comparative analysis, allowing institutions and researchers to validate whether similar engagement–performance patterns emerge across different learning contexts and analytical techniques [4], [7].

A. *Implications for Educators and Academic Institutions*

The findings have several practical implications:

Instructional guidance: Educators should encourage students to prioritize resource-based and structured content engagement, followed by reinforcement through quizzes and forum participation, rather than relying on unstructured or navigational interactions.

Early intervention: Cluster-informed monitoring enables educators to identify students who spend disproportionate time on average-impact activities, which may indicate surface-level engagement or inefficient learning strategies, and intervene with targeted feedback.

Course design: Learning activities can be redesigned to embed high-impact resources more prominently and align interactive components with core content, improving learning efficiency and assessment preparedness.

B. Limitations

Despite its contributions, this study has several limitations that should be acknowledged. First, clustering is an exploratory technique and does not establish causal relationships between engagement behaviors and assessment outcomes. While strong associations are observed, causality cannot be inferred without experimental or longitudinal intervention-based designs.

Second, the analysis relies on aggregated click-based metrics, which may not fully capture the depth of cognitive engagement or the quality of learning processes. For example, a click on a resource does not necessarily indicate comprehension, and future work could integrate temporal patterns, dwell time, or sequence-based analytics to provide richer representations of learning behavior.

Third, although OULAD provides a robust and widely used benchmark dataset, results may vary across disciplines, course structures, assessment strategies, and VLE configurations. Differences in pedagogical design or institutional context could influence which activity types are most strongly associated with performance. These limitations are consistent with those identified in broader LA and EDM literature [1], [2].

VI. Conclusion and Future Work

This paper presented an Educational Data Mining–based approach for analyzing Virtual Learning Environment (VLE) interaction logs to explain variations in student assessment outcomes. Using the Open University Learning Analytics Dataset (OULAD), a comprehensive learner–activity–assessment dataset was constructed and analyzed using K-Means clustering. The analysis identified four distinct engagement–performance clusters, of which Cluster 2 emerged as the most significant, exhibiting the highest concentration of pedagogically meaningful learning interactions and the strongest association with improved assessment performance.

A key contribution of this study is the systematic ranking of VLE activity types into most important, important, and average categories based on cluster-weighted engagement evidence. The results demonstrate that purposeful engagement with structured learning resources and core content activities plays a more influential role in academic success than overall activity volume. Activities supporting practice and interaction, such as quizzes and forums, were found to reinforce learning when combined with strong content engagement, while excessive focus on navigational or supplementary activities showed limited impact on performance. These findings reinforce the argument that engagement quality and activity semantics are critical factors in understanding online learning effectiveness.

A. Future Work

- While the findings of this study provide valuable insights, several opportunities exist for future research. First, the current analysis is based on **static, aggregated engagement measures**. Future work could incorporate **temporal modeling of engagement sequences**, enabling analysis of how learning behaviors evolve over time and how early engagement patterns influence later assessment outcomes. Sequence-based or time-aware models could offer deeper insight into learning trajectories and self-regulated learning processes.
- Second, the clustering approach adopted in this study is **exploratory**. Future research could extend this work by integrating **predictive modeling and early-warning systems** that identify at-risk learners during the course rather than after assessments are completed. Such models could build upon the identified high-impact activity categories to support proactive interventions.
- Third, although OULAD provides a robust and widely used benchmark dataset, future studies should examine **cross-course and cross-institution generalizability**. Applying the proposed methodology to additional Learning Analytics datasets and diverse disciplinary contexts would help validate the robustness of the identified activity–performance relationships and support broader adoption.

In summary, this study demonstrates the value of combining clustering-based Educational Data Mining with practical analytics tools to support data-driven teaching and learning. The proposed approach lays a foundation for more interpretable, actionable, and learner-centered analytics in higher education.

References

[1] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics and Informatics*, vol. 37, pp. 13–49, Apr. 2019, doi: 10.1016/j.tele.2019.01.007.

[2] 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, May 2024, doi: 10.18608/jla.2023.8093.

[3] S. Križanić, "Educational data mining using cluster analysis and decision tree technique: A case study," *International Journal of Engineering Business Management*, vol. 12, p. 184797902090867, Jan. 2020, doi: 10.1177/1847979020908675.

[4] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open University Learning Analytics dataset," *Scientific Data*, vol. 4, no. 1, Art. no. 170171, Nov. 2017, doi: 10.1038/sdata.2017.171.

[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, Jan. 2015, doi: 10.1007/s11528-014-0822-x.

[6] E. Kalita *et al*., “Educational data mining: A 10-year review,” *Discover Computing*, vol. 28, no. 1, Art. no. 81, May 2025, doi: 10.1007/s10791-025-09589-z.

[7] M. Hlosta, C. Herodotou, T. Papathoma, A. Gillespie, and P. Bergamin, “Predictive learning analytics in online education: A deeper understanding through explaining algorithmic errors,” *Computers and Education: Artificial Intelligence*, vol. 3, Art. no. 100108, 2022, doi: 10.1016/j.caeai.2022.100108.

[8] E. W. Black, K. Dawson, and J. Priem, “Data for free: Using LMS activity logs to measure community in online courses,” *The Internet and Higher Education*, vol. 11, no. 2, pp. 65–70, Jan. 2008, doi: 10.1016/j.iheduc.2008.03.002.

[9] A. Merceron and K. Yacef, “Educational data mining / learning analytics: Methods, tasks and current trends,” in *Proceedings of the International Conference on Educational Data Mining (EDM)*, Eindhoven, The Netherlands, 2010, pp. 1–10.

[10] P. J. Guo, J. Kim, and R. Rubin, “How video production affects student engagement: An empirical study of MOOC videos,” in *Proceedings of the First ACM Conference on Learning @ Scale*, Atlanta, GA, USA, Mar. 2014, pp. 41–50, doi: 10.1145/2556325.2566239.