Data-Driven Analysis of Learning Behaviors and Academic Success: Insights from Moodle LMS

1st Tóth Botond

ELTE Faculty of Informatics
Budapest, Hungary
tothbotond00@gmail.com

2nd dr. Kiss Attila Elemér

ELTE Faculty of Informatics
Budapest, Hungary
kiss@inf.elte.hu

***Abstract*—This project delves into the behavioral patterns of students within the Moodle LMS and their impact on academic performance. By clustering students based on engagement metrics, the study uncovers distinct learning behaviors linked to success. Predictive analytics and clustering techniques are used to identify high-performing and at-risk groups, offering recommendations for personalized support and adaptive learning interventions.**

***Index Terms*—Moodle LMS, Clustering analysis, Student engagement, Adaptive learning interventions**

I. Introduction

Online learning platforms have gained substantial traction in recent years, providing opportunities for adults to enhance their knowledge and skills. The dataset analyzed in this study was sourced from an adult learning Moodle site, "akademia.skoll.hu." Although raw data remain private to protect user privacy, a sample dataset has been made available in the accompanying repository for replication and further analysis.[1]

Understanding learner behavior is crucial for improving the effectiveness of online learning environments. Despite the growing popularity of platforms like Skillshare there is often limited analysis of the vast amount of data generated by user interactions, course completions, and assessment outcomes. Examining these behavioral patterns can provide actionable insights to improve learner engagement, optimize course design, and increase overall completion rates.

This paper focuses on exploring the patterns of participation and completion in the courses offered through the SCORM packages. By analyzing user interaction data and applying machine learning techniques such as clustering, we identify trends and offer recommendations to improve the learning experience. The subsequent sections outline the methods used, present key findings, and discuss their implications for designing more effective online courses.

II. Related Works

Numerous studies have used data from Learning Management Systems (LMS) such as Moodle to predict academic performance and analyze learning behavior. These analyses typically involve examining factors such as the frequency of logging in, the access to resources, and the patterns of interaction.

The study by [1] explores predictive modeling in Moodle LMS using machine learning techniques, emphasizing the impact of activity metrics, such as log-in frequency, on the accuracy of the final grade prediction. Similarly, [2] introduce learning analytics tools in Moodle to monitor student progress, allowing educators to identify students in need of support early in the semester.

A comparative analysis in 17 blended courses by [3] examines the correlation between LMS predictors and academic results, showcasing the scalability of data-driven analysis in Moodle. The integration of learning analytics with computational ontologies is addressed by [4], who present a taxonomy-based approach to monitor students' knowledge states within Moodle.

[5] analyze Moodle logs to assess how specific behavior patterns, like consistent engagement, correlate with academic success. The work by [6] applies machine learning on Moodle logs to detect at-risk students, providing early intervention opportunities. Similarly, [7] conduct statistical analyses of Moodle logs, focusing on activity timing and its effect on performance.

To enhance support for educators, [8] describe a Learning Analytics Dashboard for Moodle, incorporating machine learning to identify at-risk students. [9] compare two Moodle tools—Analytics Graphs and Edwiser Reports—for tracking academic progress, highlighting the efficacy of visualization in educational tracking.

Further exploration of Moodle logs in blended learning is presented by [10], who assess differences in student engagement patterns in online and in-person settings. Additionally, [11] introduce LAe-R, a Moodle plugin for assessing student performance, offering valuable insights for educators.

Beyond Moodle-specific studies, broader research on study habits, learning analytics, and academic success provides valuable context. For instance, [12] review various factors that shape study habits, including emotional and psychological aspects, while [13] emphasize the role of productive habits in academic success.

In higher education, [14] discuss how learning analytics supports academic success through data-driven feedback, a crucial aspect in monitoring LMS data. [15] investigate how study

[1]https://github.com/tothbotond00/AnalysisMoodlePlatformData

attitudes influence performance, aligning with findings from LMS-focused studies that engagement can predict academic outcomes.

Several studies focus on the predictive capability of learning analytics and educational data mining. [16] examine Moodle plugins that use learning analytics to predict student success, and [17] present insights from an empirical study on predictive analytics in Moodle, further emphasizing the application of data-driven techniques.

[18] identify Moodle activity indicators that correlate with academic performance, such as early submission and high activity levels. A similar approach is presented by [19], who propose a Moodle-based learning analytics framework for assessing course outcomes, aiding educators in adapting their strategies.

The utility of data mining in educational contexts is demonstrated by [20], who develop an ”early warning system” for identifying at-risk students in Moodle, allowing proactive intervention. Finally, [21] provide an overarching view of educational data mining and learning analytics in higher education, discussing techniques relevant to LMS data analysis for predicting academic success.

Collectively, these studies highlight the importance of LMS data and learning behavior in predicting and improving academic performance. Activities such as timely assignment submission, consistent resource access, and proactive engagement have been shown to be significant predictors of student success.

III. Analysis of Course Popularity and Completion Patterns

A. Overview

This section analyzes the interactions and completion patterns of courses offered through the SCORM packages. Two visualizations were used to explore these trends: a bar graph illustrating the total interactions and completions for each course, and a clustered scatter plot to identify patterns in course participation.

B. Interactions and Completions by Course

The bar chart (Figure 3) compares the total interactions and completions for various courses. Key observations include:

- Certain courses exhibit significantly higher total interactions, indicating their popularity among users.
- High interaction courses generally demonstrate a proportional increase in total completions, suggesting engagement as a critical factor in course completion.
- A subset of courses displays lower completions relative to their interactions, pointing to potential issues such as user drop-off or challenges in course content.

These findings highlight the importance of designing engaging and accessible course content to improve user retention and completion rates.

C. Clustering Analysis of Course Engagement

To further explore the data, a clustering analysis using the K-Means algorithm was conducted. Courses were clustered based on features such as total interactions, total completions, average grade percentage, and completion rate. The resulting scatter plot (Figure 4) reveals three distinct clusters:

- **Cluster 0:** Courses with moderate interactions and relatively higher completion rates. These courses appear well-designed and effective in engaging learners.
- **Cluster 1:** Courses with lower interactions and completion rates. These courses may require content revisions or additional promotion to improve engagement.
- **Cluster 2:** Courses with high interactions and high completion rates. These courses likely represent the most popular and successful offerings.

D. Implications and Recommendations

This analysis provides actionable insights into course performance:

- Courses with high interactions and completions (Cluster 2) can serve as benchmarks for designing future content.
- Courses with high interactions but lower completions warrant closer examination to identify and mitigate user drop-offs.
- Less popular courses (Cluster 1) may benefit from targeted improvements in content design, delivery mechanisms, or marketing strategies.

This clustering approach enables the identification of patterns in user behavior, providing valuable insights for optimizing course offerings.

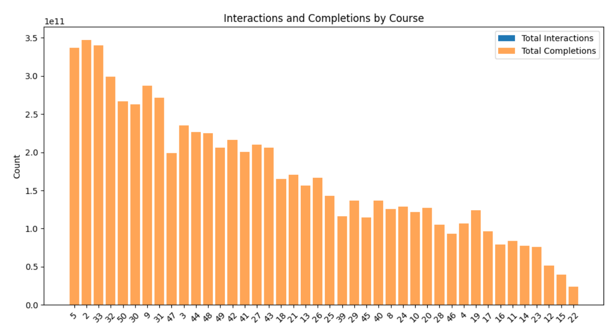

Fig. 1. Total Interactions and Completions by Course

IV. Clustering-Based Analysis of Academic Performance Using Education and Location Features

A. Optimal Clustering and Dataset Insights

The dataset sourced from an adult-learning Moodle platform was analyzed using clustering techniques to uncover patterns in learner performance. The features used for clustering included the calculated grade percentage (based on module grades), categorical education levels, and geographic location information.

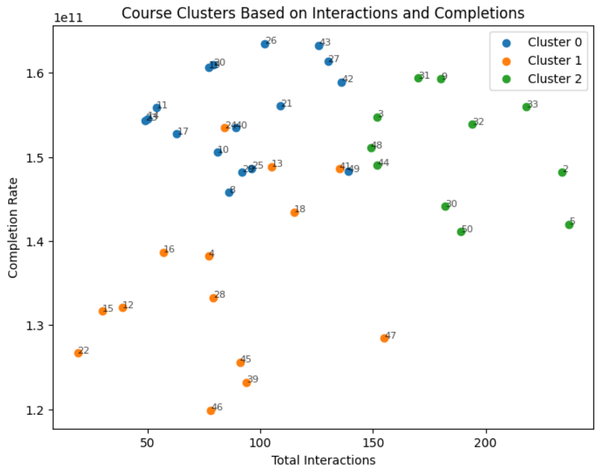

Fig. 2. Course Clusters Based on Interactions and Completions

To determine the appropriate number of clusters, the Elbow Method was employed (Figure **??**). The method revealed a sharp decrease in inertia up to $k = 2$, followed by a gradual decline. Thus, $k = 2$ was selected as the optimal number of clusters for further analysis. These clusters represent distinct learner profiles with significant variations in performance.

B. Cluster Characteristics and Observations

The analysis produced two clusters that distinguish learners primarily based on their grade percentages:

- **Cluster 0: High Performers**
  - Average grade percentage: **99.99%**.
  - Primarily composed of learners with a university-level education or equivalent (40.06% from ”egyetem”).
  - A significant proportion of users in this cluster are from urban areas, particularly Budapest, which represents 12.39% of the cluster.
  - Explanation: The high performance observed in this cluster can potentially be attributed to better access to resources, such as high-quality learning materials, reliable internet access, and academic support in urban areas. University-level education might also correlate with stronger learning habits or prior academic preparation.
- **Cluster 1: Moderate Performers**
  - Average grade percentage: **90.30%**.
  - More diverse educational backgrounds, including a larger representation of vocational schools (”szakiskola,” 17.45%) and high schools (”gimnázium,” 15.71%).
  - Includes learners from a wider variety of locations, with a smaller concentration from Budapest (less than 10%).
  - Explanation: The lower performance in this cluster may be influenced by limited access to learning resources or less formal academic training. Learners in this group may also balance other responsibilities, such as employment, which could impact their engagement levels.

C. Visualization of Clusters

To validate the clustering results, a two-dimensional visualization was created using Principal Component Analysis (PCA), as shown in Figure **??**. The PCA projection illustrates a clear separation between the two clusters:

- Cluster 0 (high performers) forms a distinct grouping, suggesting consistent learner characteristics within this cluster.
- Cluster 1 (moderate performers) displays more variation, which aligns with the diversity observed in education levels and locations.

These visualizations support the hypothesis that educational background and geographic location are significant factors influencing academic performance on the platform.

D. Insights and Possible Explanations

The clustering analysis offers several insights into the behavior and performance of learners:

- **Impact of Urban Environments:** The high proportion of high-performing learners in Budapest may reflect the availability of better infrastructure, educational resources, and internet connectivity in urban areas.
- **Role of Education Levels:** The strong presence of university-educated learners in Cluster 0 highlights the potential correlation between formal education and successful course completion. University students may also be more accustomed to self-directed learning, a critical skill in online education.
- **Diverse Profiles in Moderate Performers:** Cluster 1 reveals a more heterogeneous mix of learners, including those from rural areas and non-traditional educational backgrounds. This diversity suggests that additional support, such as personalized feedback or tailored learning resources, could help improve outcomes for this group.

E. Implications for Platform Design

The findings have significant implications for the design of online learning platforms:

- **Targeted Support for Moderate Performers:** To address the needs of Cluster 1 learners, platforms could implement features such as adaptive learning paths, reminders for course progress, and more interactive elements to maintain engagement.
- **Enhancing Urban Reach:** While urban learners already exhibit high performance, extending access to advanced tools (e.g., AI-driven tutoring) could further boost their outcomes and set benchmarks for other regions.
- **Bridging the Digital Divide:** For rural and vocational learners, initiatives such as offline learning resources, mobile-optimized content, and flexible course schedules could help close the performance gap.

F. *Limitations and Future Directions*

While clustering reveals significant trends, there are limitations to this analysis. For example, the use of a limited feature set (grade percentages, education levels, and location) may overlook other critical factors such as time spent on tasks, forum participation, or motivational aspects. Future research could expand the feature set and employ more advanced clustering techniques to uncover additional insights.

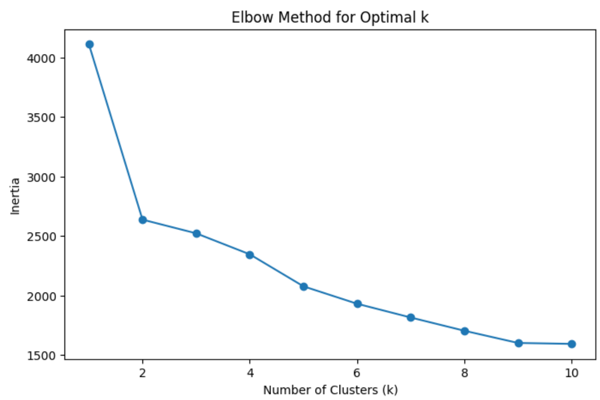

Fig. 3. Elbow Method indicating optimal clusters ($k = 2$)

Fig. 4. Cluster visualization using Principal Component Analysis (PCA).

V. Temporal Analysis of User Activity

A. *Objective*

This section investigates temporal trends in user activity on the Moodle platform, including weekly engagement patterns, seasonality, and hourly activity distributions. Using timestamped data, we explore fluctuations in user interactions and project future trends to support platform optimization efforts.

B. *Methodology*

- **Preprocessing:** Unix timestamps from columns such as *CourseModuleCompletion* and *CourseCompletionStart* were converted into human-readable datetime formats. Null or zero timestamps (e.g., 1970-01-01) were filtered out to ensure accuracy. A unified *activity_date* column was created, prioritizing completion dates and falling back to start dates when necessary.
- **Weekly Activity Trends:** Activity counts were aggregated on a weekly basis to observe overall participation trends. Missing dates were filled with zero activity to create a continuous timeline. Figure 5 shows the weekly user activity over time, highlighting periods of high engagement and significant drops in participation.
- **Seasonality Analysis:** A seasonal decomposition of the weekly activity data was conducted to identify underlying trends, seasonal patterns, and residual fluctuations. The decomposition (Figure 6) reveals recurring cycles and deviations, which are key to understanding user behavior.
- **Hourly and Daily Patterns:** Hourly and day-of-week activity patterns were analyzed to identify engagement peaks during specific time periods. A heatmap (Figure 7) visualizes the distribution of activity, highlighting higher participation rates on Mondays and during late morning hours.
- **Future Activity Forecasting:** Using Prophet, a time-series forecasting model, weekly activity was projected for 12 weeks into the future. Figure 8 shows the forecasted trends with confidence intervals, providing insights into potential future participation levels.

C. *Results and Insights*

- **Engagement Trends:** Weekly user activity demonstrates notable fluctuations, with a sharp decline in mid-year participation, followed by periods of moderate recovery.
- **Seasonality:** The seasonal decomposition (Figure 6) highlights recurring engagement patterns, reflecting cyclical user behavior, likely driven by course schedules or external events.
- **Hourly Engagement:** As depicted in the heatmap (Figure 7), Monday mornings show the highest activity levels, suggesting that learners engage more at the start of the week.
- **Future Trends:** The forecast model (Figure 8) predicts a gradual decline in activity over the next 12 weeks, with wide variability reflecting potential external influences.

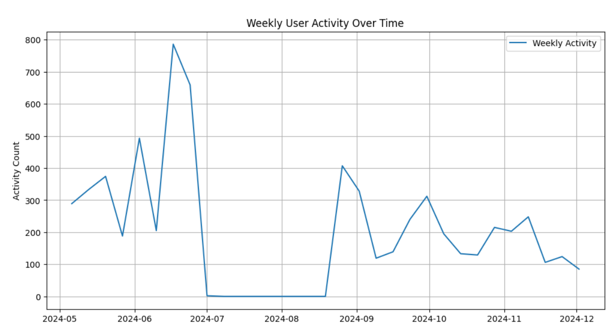

Fig. 5. Weekly User Activity Over Time

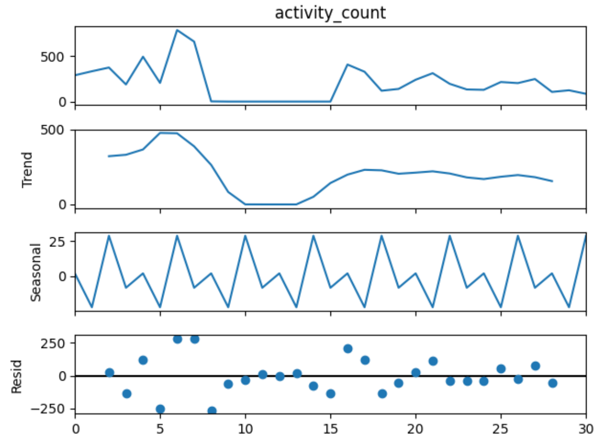

Fig. 6. Seasonal Decomposition of Weekly Activity

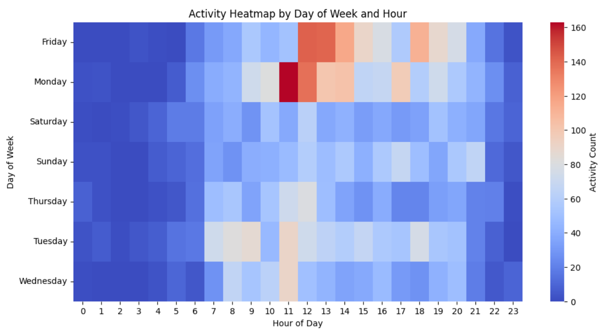

Fig. 7. Activity Heatmap by Day of Week and Hour

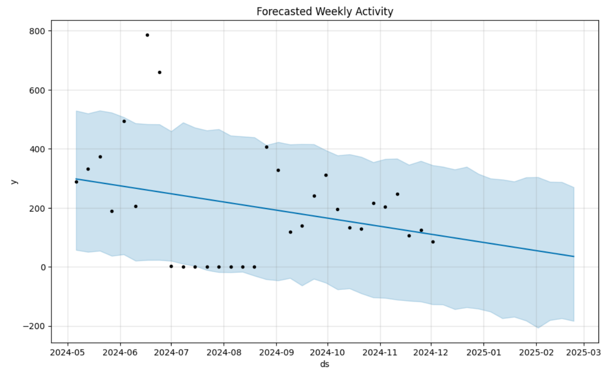

Fig. 8. Forecasted Weekly Activity

VI. Future Works

While this study provides valuable insights into learner behavior and academic performance on an online learning platform, there are several avenues for further research and development to build upon these findings.

A. Integration of Advanced Machine Learning Techniques

Future research could incorporate more advanced machine learning algorithms, such as hierarchical clustering, neural networks, or ensemble models, to capture more complex patterns in learner behavior. Predictive modeling, for instance, could be used to forecast course completion rates or identify at-risk learners early in their learning journeys.

B. Temporal Analysis of Learning Behaviors

A temporal component could be added to the analysis by examining how learning behaviors evolve over time. This might involve tracking individual engagement levels, completion rates, and grade progression across multiple courses or semesters, providing insights into long-term learning patterns.

C. Personalized Recommendations

Leveraging the insights from this dataset, future studies could explore the development of personalized learning recommendations. By identifying clusters of learners with similar characteristics, the platform could suggest tailored learning paths, resource recommendations, or adaptive content delivery to enhance individual learning outcomes.

D. Expanded Feature Set

Future work should include a broader set of features, such as detailed interaction logs, forum activity, and peer interactions. This could provide a more holistic view of learner behavior and allow for more nuanced analyses of factors influencing academic success.

E. Cross-Platform and Demographic Comparisons

Comparative studies across different online learning platforms or demographic groups would help generalize the findings. For instance, analyzing how factors like socioeconomic status, device usage, or regional disparities affect learning behaviors could provide valuable cross-context insights.

F. Behavioral and Psychological Factors

The integration of behavioral and psychological factors, such as motivation, self-regulation, or satisfaction levels, into the analysis would offer deeper insights into the drivers of learner engagement and performance. Surveys or feedback mechanisms could be used to enrich the dataset with qualitative data.

G. Real-Time Monitoring and Interventions

Building a real-time monitoring system based on the findings of this study could enable early interventions. For example, identifying learners who show signs of disengagement or struggle and offering timely support could improve overall course retention and completion rates.

H. Broader Scope of Educational Contexts

Finally, the methodologies and findings from this dataset could be extended to other educational contexts, such as corporate training platforms, K-12 education, or university e-learning systems. This would further validate the scalability and applicability of the analytical framework presented in this paper.

By addressing these directions, future research can contribute to creating more engaging, inclusive, and effective online learning environments while continuing to advance the field of educational data mining and learning analytics.

References

[1] Mia Mužinić, Antonela Sikavica, Petra Zelić, Ani Grubišić, and Ines Šarić Grgić. Predictive modeling of student performance in moodle lms using learning analytics. In *2024 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–7, 2024.

[2] Cristina Alonso-Fernández, José L. Jorro-Aragoneses, Carlos M. Alaíz, and Pilar Rodríguez. Learning analytics tools to analyze progress and results with moodle lms data. In *2024 IEEE Global Engineering Education Conference (EDUCON)*, pages 01–07, 2024.

[3] Rianne Conijn, Chris Snijders, Ad Kleingeld, and Uwe Matzat. Predicting student performance from lms data: A comparison of 17 blended courses using moodle lms. *IEEE Transactions on Learning Technologies*, 10(1):17–29, 2017.

[4] Laécio A. Costa, Laís N. Salvador, and Ricardo R. Amorim. Evaluation of academic performance based on learning analytics and ontology: a systematic mapping study. In *2018 IEEE Frontiers in Education Conference (FIE)*, pages 1–5, 2018.

[5] Martín Liz-Domínguez, Martín Llamas-Nistal, Manuel Caeiro-Rodríguez, and Fernando Mikic-Fonte. Lms logs and student performance: The influence of retaking a course. In *2022 IEEE Global Engineering Education Conference (EDUCON)*, pages 1970–1974, 2022.

[6] Mariela Mizota Tamada, Rafael Giusti, and José Francisco de Magalhães Netto. Predicting student performance based on logs in moodle lms. In *2021 IEEE Frontiers in Education Conference (FIE)*, pages 1–8, 2021.

[7] Xiangfeng Tan, Jinhua She, Shumei Chen, Sumio Ohno, and Hiroyuki Kameda. Analysis of student learning behavior based on moodle log data. In *2021 14th International Conference on Human System Interaction (HSI)*, pages 1–4, 2021.

[8] Cristian Cechinel, Mateus De Freitas Dos Santos, Caio Barrozo, Jesiel Emerim Schardosim, Eduardo de Vila, Vinicius Ramos, Tiago Primo, Roberto Munoz, and Emanuel Marques Queiroga. A learning analytics dashboard for moodle: Implementing machine learning techniques to early detect students at risk of failure. In *2021 XVI Latin American Conference on Learning Technologies (LACLO)*, pages 130–136, 2021.

[9] Joe Llerena, Nohely Álava Morán, and Jonathan Zamora-Galindo. Learning analytics for student academic tracking, a comparison between analytics graphs and edwiser reports. In *2021 Second International Conference on Information Systems and Software Technologies (ICI2ST)*, pages 101–107, 2021.

[10] Adrián Pérez Suay, Steven Van Vaerenbergh, Maria Piles, Valero Laparra, Ana B. Pascual-Venteo, Ana B. Ruescas, Roberto Fernandez-Moran, Marina Martínez-García, Jose E. Adsuara, Julia Amorós, Jordi Muñoz-Marí, Irene Epifanio, and Miguel-Ángel Fernández-Torres. Learning about student performance from moodle logs in a higher education context. In *2022 XII International Conference on Virtual Campus (JICV)*, pages 1–4, 2022.

[11] Ourania Petropoulou, Katerina Kasimatis, I. Dimopoulos, and Symeon Retalis. Lae-r: A new learning analytics tool in moodle for assessing students' performance. *Bulletin of the Technical Committee on Learning Technology*, 16:2–5, 01 2014.

[12] Ahmad L. Rafiq. Study habits and academic performance among students: A systematic review. *Research Review Journals*, 10(5):132–135, 2021.

[13] Angela L. Duckworth and Brian M. Galla. The science of habit and its implications for student learning and well-being. *Educational Psychology Review*, 32:465–487, 2020.

[14] Dirk Ifenthaler and Jessica Y. Yau. Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 68(4):1961–1990, 2020.

[15] Jhoselle D. Tus. The influence of study attitudes and study habits on the academic performance of the students. *International Journal of Research in Educational Science*, 6:52–60, 2020.

[16] Zhang Yaqun, Ghandour Ahmad, and Shestak Viktor. Using learning analytics to predict students performance in moodle lms. *ResearchGate*, 2020.

[17] Duch Dynil, May Madeth, and George Sébastien. Enhancing predictive analytics for students' performance in moodle: Insight from an empirical study. *Journal of Data Science and Intelligent Systems*, 2021.

[18] Ewa Młynarska, Derek Greene, and Cunningham Pádraig. Indicators of good student performance in moodle activity data. *arXiv preprint arXiv:1601.02975*, 2016.

[19] Yassine Sahar, Kadry Seifedine, and Sicilia KadryMiguel-Angel. A framework for learning analytics in moodle for assessing course outcomes. In *International Conference on Education and New Learning Technologies*, 2016.

[20] Min Min, Junlin Cheng, and Yan Zhang. Research on learning early warning and intervention for english major undergraduates based on moodle learning management system (lms). In *Proceedings of the 2024 Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Education Digitalization and Computer Science*, EDCS '24, page 93–97, New York, NY, USA, 2024. Association for Computing Machinery.

[21] George Siemens and Ryan Baker. Educational data mining and learning analytics in higher education. *British Journal of Educational Technology*, 43(3):465–475, 2012.