



Article

Digital Footprints of Academic Success: An Empirical Analysis of Moodle Logs and Traditional Factors for Student Performance

Dalia Abdulkareem Shafiq^{1,*}, Mohsen Marjani¹, Riyaz Ahamed Ariyaluran Habeeb², and David Asirvatham¹

¹ School of Computer Science, Faculty of Innovation and Technology, Taylor's University, Subang Jaya 47500, Malaysia; mohsen.marjani@taylors.edu.my (M.M.); david.asirvatham@taylors.edu.my (D.A.)

² Department of Information System, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia; riyaz@um.edu.my

* Correspondence: daliaabdulkareemshafiq@sd.taylors.edu.my

Abstract: With the wide adoption of Learning Management Systems (LMSs) in educational institutions, ample data have become available demonstrating students' online behavior. Digital traces are widely applicable in Learning Analytics (LA). This study aims to explore and extract behavioral features from Moodle logs and examine their effect on undergraduate students' performance. Additionally, traditional factors such as demographics, academic history, family background, and attendance data were examined, highlighting the prominent features that affect student performance. From January to April 2019, a total of 64,231 students' Moodle logs were collected from a private university in Malaysia for analyzing students' behavior. Exploratory Data Analysis, correlation, statistical tests, and post hoc analysis were conducted. This study reveals that age is found to be inversely correlated with student performance. Tutorial attendance and parents' occupations play a crucial role in students' performance. Additionally, it was found that online engagement during the weekend and nighttime positively correlates with academic performance, representing a 10% relative increase in the student's exam score. Ultimately, it was found that course views, forum creation, overall assignment interaction, and time spent on the platform were among the top LMS variables that showed a statistically significant difference between successful and failed students. In the future, clustering analysis can be performed in order to reveal heterogeneous groups of students along with specific course-content-based logs.



Academic Editor: Han Reichgelt

Received: 5 August 2024

Revised: 29 September 2024

Accepted: 9 October 2024

Published: 28 February 2025

Citation: Abdulkareem Shafiq, D., Marjani, M., Ahamed Ariyaluran Habeeb, R., & Asirvatham, D. (2025). Digital Footprints of Academic Success: An Empirical Analysis of Moodle Logs and Traditional Factors for Student Performance. *Education Sciences*, 15(3), 304. <https://doi.org/10.3390/educsci15030304>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Moodle; online learning; student performance; learning management system; learning analytics; statistical tests

1. Introduction

Students interact with a university through various activities such as portal login, uploading assignments, and class attendance, among many others, and these actions often leave behind digital traces that can be used to analyze how the university is managing and providing help for students. Such traces can be stored as potentially widely applicable data in Learning Analytics (LA). It is the process of analyzing, measuring, collating, and reporting data to make decisions about the progress made by students and educators. LA is emerging as a top priority for many institutions (Gonzalez-Nucamendi et al., 2022), especially leveraged during world crises such as COVID-19.

Learning Analytics (LA) and the scope of predicting students' performance is multidimensional (Waheed et al., 2020) and, therefore, can be tackled through various approaches from multiple perspectives to explore and analyze the problem, including identifying

at-risk students, understanding the factors affecting the student's performance, and the early prediction of dropouts and withdrawals during or after a course (Shafiq et al., 2022). Thus, choosing the right approach highly depends on the prediction goal and the dataset structure. In this study, LA is demonstrated by extracting students' behavior and interaction features using log files from Learning Management Systems (LMSs) to analyze student performance.

The study of the relationship between behavioral interactions and academic performance in courses supported by Virtual Learning Environments (VLEs) has been the focus of LA (Agudo-Peregrina et al., 2014; Wong & Li, 2020). This aspect has emerged as the most crucial for planning and implementing effective learning processes. Researchers have used physiological sensors and human observers to collect data on learners' behavior in Virtual Learning Environments (VLEs) (Aguagallo et al., 2023; Prakash, 2023). However, data collection using human observers is labor-intensive and may be biased (Llerena-Izquierdo et al., 2023). To overcome these challenges, researchers have developed mechanisms to automatically collect real-time data on learners' interaction behavior in online learning (Khor & Dave, 2022).

Data collection mechanisms, such as logs, open up new opportunities for data collection and provide insights into learners' behavior and learning processes in VLEs. By applying AI and ML tools to analyze the collected data, researchers aim to understand the behavioral patterns contributing to higher learning gains and predicting learning outcomes (Lavidas et al., 2022). This information can be used to provide scaffolds and adaptive feedback to enhance the learning experience in VLEs.

The platform Modular Object-Oriented Dynamic Learning Environment (Moodle) has long been identified as the most popular and preferred open-source LMS (Moodle, 2024). It has been known as an advanced learning platform used in multiple disciplines, particularly in STEM education. It is mainly used within university settings, focusing on undergraduate studies. The platform has effectively improved student performance, satisfaction, and engagement while enhancing flexibility in learning environments (Gamage et al., 2022; Furqon et al., 2023). Using LMSs for obtaining digital traces to analyze student behavior online eliminates the need for time-consuming data collection methods. However, LMS log data are raw and may not provide tangible measurements, necessitating investigations of how and whether such data can be utilized for Learning Analytics.

The analysis of the effect of Learning Management Systems (LMSs) on student performance in different educational settings has gained attention among researchers, and it is still ongoing (İşikgöz, 2024; Khairy et al., 2024). Studies have shown that utilizing LMSs has a beneficial effect on academic performance among students and fosters a favorable perception of LMS implementation in educational endeavors (Furqon et al., 2023). Additionally, using LMSs allows for the analysis of student trajectories and identifying factors that lead to successful or unsuccessful learning outcomes (Shaimov et al., 2022). Research conducted in the Computer Science Department of a Nigerian institution found that using LMSs improved students' academic performance and awareness of collaborative and academic study (Ifeanyi & Chinonso, 2023). Furthermore, statistically based approaches using data from the Moodle LMS have indicated that the number of interactions with the LMS can serve as a good indicator of student performance and can be used to track their learning progress (Suay et al., 2022). On the other hand, analysis of log-based profiling and behavior on LMSs revealed that an online learner with fewer interactions could drop out or achieve lower academic performance (Çakiroğlu et al., 2024).

Moreover, given that the nature of student performance is multifactorial and complex, relying on limited academic and LMS factors may not capture the entire picture of why

students perform in a certain way, and their performance can be heavily influenced by other factors such as socio-economic background and financial factors, among many.

In a blended learning environment setting, which is the focus of this paper, the integration of traditional factors alongside those derived from LMSs is necessary to provide a comprehensive understanding of student success and informs targeted interventions. Factors such as age, gender, and socioeconomic status significantly influence academic performance, as highlighted in studies that incorporate these variables into performance assessments (Loan et al., 2024). Other traditional factors include Nationality, where research indicates that incorporating such factors is crucial for understanding disparities in educational outcomes (Sommerville & Singaram, 2018; Banda et al., 2023). Fernández-Leyva et al. (2021) found that such factors significantly influence academic success, particularly among immigrant students, where cultural and social skills play a pivotal role.

Furthermore, several studies have shown a positive correlation between class attendance and academic performance (Jiang, 2022; Ng et al., 2022; Latif Khan et al., 2019). In one study, lecture class attendance was found to have a positive correlation with final exam results. Another study used a random forest algorithm to identify the most important independent variables affecting student performance and found that the number of school absences was one of the critical factors. A data-mining approach also revealed that absenteeism was a significant factor in predicting student performance. Furthermore, research conducted on undergraduate medical students showed that attendance had a significant impact on their academic achievements. Overall, these findings suggest that attendance and leave of absence play a crucial role in determining student performance. However, these studies did not consider the different types of attendance in class, particularly in a university learning environment, where class can be conducted in the form of lectures, tutorials, and practical assignments, which may reveal interesting insights for a more tailored intervention approach.

While traditional metrics are essential, the growing reliance on online learning necessitates a balanced approach that incorporates both sets of factors to effectively monitor and enhance student performance. For example, academic results such as marks or grades are highly important to predict the performance of the student; however, these are still considered insufficient indicators of why the student is performing in this way or at that level. Therefore, these are suggested to be used in conjunction with online metrics to provide a holistic view of performance (Shou et al., 2024), potentially reducing the misinterpretations of student capabilities.

1.1. Research Questions

This study addresses the importance of understanding how student interactions within the Moodle platform impact their academic performance by examining both the log variables (such as the total logins, time spent on the platform, interaction with different activities, and hourly login patterns) and traditional factors (such as demographics, academic history, family background, and attendance data). This research aims to uncover insights into the relationship between online engagement and academic success by addressing four (4) key research questions formulated below:

- RQ1. How do traditional features such as the student's demographic profile, academic history, and family background correlate with the final academic mark?
- RQ2. What attendance features are significantly associated with academic performance?
- RQ3. How do different patterns of LMS login and engagement contribute to academic success or failure among undergraduate students?
- RQ4. Which features of the LMS demonstrate the strongest correlation with academic performance among undergraduate students?

1.2. Research Highlight and Contribution

This study is substantial, as it sheds light on how online interactions in blended educational settings impact student performance. It delivers constructive information for educators to improve their teaching strategies and use of LMSs. Additionally, it was found that most of the studies in the literature focus on overall behavior patterns in LMSs, ignoring other factors, such as temporal distribution (Li et al., 2022), that can capture distinct patterns and students' behavior. Therefore, we aim to analyze the impact of such features on students' performance, potentially contributing to the ongoing evolution of online education tools. Furthermore, as for attendance, this feature may include different attendance types in higher education, such as lectures or tutorials, each of which can reveal interesting findings catered to each type and may call for specific interventions by institutions; however, in the literature, researchers often use overall attendance only (Shafiq et al., 2022; Li et al., 2022).

Our contribution to the literature can be summarized as follows:

- Establish robust correlations using rigorous statistical methods and heatmaps.
- Analyze a taxonomy of factors and available features in universities that can capture undergraduate student performance.
- Provide valuable insights and investigate students' online learning behavior.
- Identify key factors that are significant determinants of academic performance, aiding future researchers in the feature selection process for accurate predictive models, as well as institutions by identifying valuable markers for intervention and support strategies.
- Expand on the existing literature by examining temporal online features and engagement analysis of LMS patterns—an area that has received less attention.
- This study used a new dataset from a private university in Malaysia, making it particularly beneficial for similar institutions.

The rest of the paper is organized as follows. Section 2 explains the methodology adopted for the analysis, highlighting the data collection, pre-processing, feature extraction, tools, and statistical tests used. Section 3 includes the results and key findings of the tests. Section 4 consists of the discussion and implications by relating the findings with the existing literature. In Section 5, our research concludes by summarizing the importance and content of this study paper, the limitations, and, finally, future work.

2. Materials and Methods

This section explains the sources of the dataset, pre-processing steps, feature extraction process, tools, and the analysis tests used.

2.1. Data Collection

The data used in this study were obtained from a private university in Malaysia (the university name is masked to preserve anonymity and confidentiality) through the Microsoft Access (.accdb) database, consisting of 8 tables, 4 of which were exported into Excel (.xlsx) files to be used for feature extraction and further analysis in Jupyter Notebook, using the Python 3 programming language. Behavioral online data about students were collected from the Moodle LMS, which consisted of 19 blended modules taught from January to April 2019 in the School of Computing and IT.

Since the objective of this study is to utilize online features extracted from LMS logs along with traditional records and exam results, the final exam mark was used as the dependent variable in most cases for the analysis of academic performance. Consequently, only students who attempted the final exam and whose mark was recorded were included in this study. The final sample included 221 students from 5 undergraduate courses. A summary

of the courses is depicted in Table 1, where Computer Security and Forensics and Internet Technologies are sub-majors of Computer Science and Internet Technologies, respectively.

Table 1. Courses and sample size.

No.	Course Description	Sample Size
1	Software Engineering	89
2	Computer Science	83
3	Information and Technology	14
4	Computer Security and Forensics	25
5	Internet Technologies	10

Table 2 summarizes the subjects taken by students and included in this study. In this case, the same student could enroll in up to two or three modules per semester. Additionally, modules with only one student typically indicated that the student was a Computing major, although students from different majors could also enroll. Therefore, the number of students in a module does not represent the total number of students who took that subject.

Table 2. Number of students according to the modules (subjects).

No.	Subject Description	Number of Students
1	Entrepreneurship and Small Business	5
2	Understanding Entrepreneurialism	55
3	Organizational Communication	22
4	Professional Computing Practice	49
5	Event Marketing	38
6	Introduction to Finance	60
7	Information Design	15
8	Systems Fundamentals	1
9	World Languages	9
10	Chinese Language 1	19
11	Korean 1	4
12	Introduction to Management	45
13	Principles of Marketing	16
14	Hubungan Etnik	79
15	Tamadun Islam dan Tamadun Asia	79
16	Bahasa Melayu Komunikasi 2	22
17	Pengajian Malaysia 3	21
18	Bahasa Kebangsaan A	1
19	Malaysian Indigenous Cultures	1

Next to LMS data, demographic profiles, academic history, and family background were included to attain a holistic view of student performance. Lastly, assessment data such as attendance and final exam results were collected. Attendance involves physical participation in classes. This feature is represented by different attendance types, such as lectures and tutorials, each of which can reveal interesting findings catered for each type and may call for specific interventions by institutions.

2.2. Data Pre-Processing and Feature Extraction

Data pre-processing, including Data Transformation, Imputation ([Abu Zohair, 2019](#)), Label Encoding ([Abu Zohair, 2019](#)), and Conditional Aggregation ([Al-Sulami et al., 2023](#)), was performed prior to the selection of features to transform raw data ([Viswanathan et al., 2021](#)) into a format suitable for statistical analysis.

Feature extraction is a subprocess of feature engineering; this involves producing new features from the available raw data based on the problem statement (Mubarak et al., 2021). This process reduces the dimensionality of the initial set containing the raw data to be more manageable and easy to interpret. In this study, this process was used to create new attributes to address the objectives of the analysis.

LMS log files from Moodle were used to gather online features, as they were recognized as the most practical data source in terms of the level of information and the time and labor intensity of coding (You, 2016). Typically, user log data are not stored orderly and may not directly present meaningful information on students' behaviors (Arizmendi et al., 2023). Generally, the user log file consists of a timestamp and log ID, along with a description of the actions taken. A snippet of the LMS log data used in this study is demonstrated in Figure 1. However, these data should not be directly used as input for modeling, as they need to be converted to specific data types and relevant features to make accurate theoretical and predictive conclusions. The LMS structure consists of many components, each categorized for particular activities. A summary of the components considered in this study for feature extraction is shown through a taxonomy in Figure 2.

log ID	subject id	subject name	component	target	action				
1 MGT60104T0558-M		Introduction to Management	core	course					viewed
2 MGT60104T0558-M		Introduction to Management	mod_quiz	course_module					viewed
3 MGT60104T0558-M		Introduction to Management	core	course					viewed
4 MGT60104T0558-M		Introduction to Management	mod_resource	course_module					viewed
5 MGT60104T0558-M		Introduction to Management	mod_resource	course_module					viewed
eventname	datecreated	Student ID	status	intake	Semester	intake start	intake end	intake end 2	WEEKS01
corecourseviewed	2019-01-01 13:08:45	A	Active	201703	7	2017-03-22 00:00:00	2020-03-27 00:00:00	2020-03-27	-13
mod_quizcourse_moduleviewed	2019-01-01 13:08:48	A	Active	201703	7	2017-03-22 00:00:00	2020-03-27 00:00:00	2020-03-27	-13
corecourseviewed	2019-01-01 21:12:26	B	Active	201708	6	2017-08-23 00:00:00	2020-09-04 00:00:00	2020-09-04	-13
mod_resourcecourse_moduleviewed	2019-01-01 21:13:30	B	Active	201708	6	2017-08-23 00:00:00	2020-09-04 00:00:00	2020-09-04	-13
mod_resourcecourse_moduleviewed	2019-01-01 21:13:43	B	Active	201708	6	2017-08-23 00:00:00	2020-09-04 00:00:00	2020-09-04	-13

Figure 1. Snippet of LMS log data.

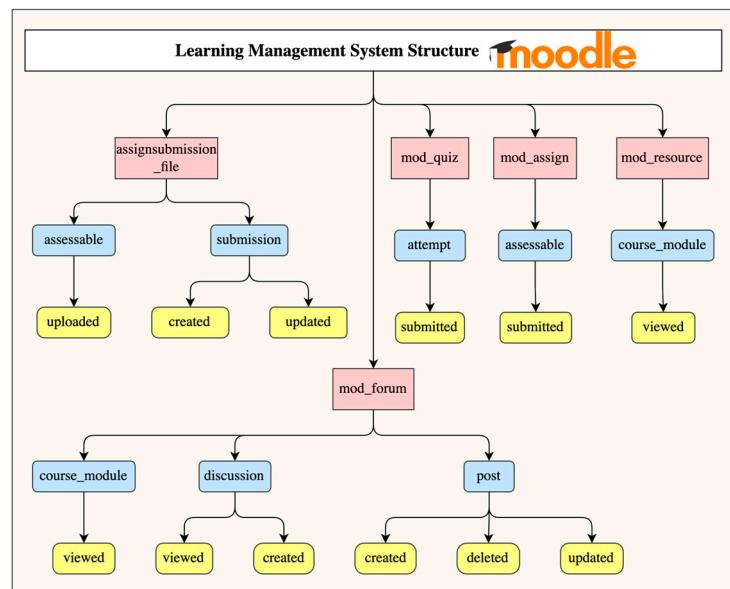


Figure 2. A taxonomy of Moodle components, targets, and actions.

Therefore, Moodle's report structure was unsuitable for Learning Analytics and required a transformation process. Data obtained from the Moodle platform, a platform students use to access their modules online, were used to extract online features such as LMS interaction. This consisted of a total of 64,231 instances, known as log entries. Such logs represented the activities and interactions of students with the modules taken. Click-stream data have been widely used among researchers in the LA field to predict academic performance, as they are often associated with a student's online behavior (Shafiq et al.,

2022; Alnasyan et al., 2024). A total of 13 features were extracted from the LMS, as depicted in Table 3. They were grouped into three categories: firstly, “Login Behaviour”, which captures the frequency and different daily timing of logins, indicating how students interact with the online platform; secondly, “Engagement Behaviour”, which reflects student engagement with specific course components such as modules, assignments, and forums; and lastly, “Time Spent”, which represents the total time spent by students, being a possible indicator of their study effort. However, few studies in the literature have focused on the temporal distribution of login behavior, for example, on patterns of students’ logins during different days of the week and hours of the day; therefore, this study delves deeper into understanding how these features affect students’ marks.

Table 3. Summary of extracted online (LMS) features.

No.	New Attribute	Description	References
1	Total logins ¹	Number of times student logs in the platform	(Bainbridge et al., 2018; Bravo-Agapito et al., 2021; Cui et al., 2020; Mwalumbwe & Mtebe, 2017; Kondo et al., 2017)
2	Weekday login ¹	Number of logins by User during the week (Monday–Friday)	(Cui et al., 2020)
3	Weekend login ¹	Number of logins by User during the weekend (Saturday–Sunday)	
4	earlyAM ¹	Number of logins by student between 12:00:00 a.m.–5:59:59 a.m.	(Quinn & Gray, 2019)
5	lateAM ¹	Number of logins by student between 5:59:59 a.m.–11:59:59 a.m.	
6	latePM ¹	Number of logins by student between 11:59:59 p.m.–7:59:59 p.m.	
7	lateAM ¹	Number of logins by student between 7:59:59 p.m.–11:59:59 p.m.	
8	Module view ²	Number of course/module views by student	Akçapınar et al., 2019; Helal et al., 2018)
9	Assign interaction ²	Number of total interactions with assignment by student	(Bravo-Agapito et al., 2021; Cui et al., 2020)
10	Assign submit ²	Number of assignment submission by student	(Bravo-Agapito et al., 2021; Kondo et al., 2017)
11	Assign view ²	Number of assignment views by student	(Bainbridge et al., 2018; Bravo-Agapito et al., 2021; Cui et al., 2020; Helal et al., 2018)
12	Forum view ²	Number of forum discussion views by student	(Bainbridge et al., 2018; Cui et al., 2020; Akçapınar et al., 2019)
13	Forum create ²	Number of forum discussion created by student	(Mwalumbwe & Mtebe, 2017; Kondo et al., 2017; Akçapınar et al., 2019)
14	Timespent ²	Average total time spent daily by student in minutes	(Bainbridge et al., 2018; Bravo-Agapito et al., 2021; Cui et al., 2020; Mwalumbwe & Mtebe, 2017; Kondo et al., 2017)

All features refer to engagement, whereas ¹ refers to features related to clickstream, and ² refers to academic engagement.

Furthermore, traditional features were extracted based on their availability and consistency in the dataset from tables such as Student Profiles, Attendance, and Exam Records. These features were selected based on the extensive existing literature and according to the interest of the analysis objective. A summary of the extracted features is demonstrated in Table 4.

Records were collected for each class type to extract data related to class attendance, such as lectures and tutorials. It was essential to gather information pertaining to bachelor students’ class performances in computing courses. It was found that attendance count and whether the student took a leave of absence were the strongest predictors of their performance; therefore, these features could be created using raw data. Using columns such as the student ID, attendance type, and attendance (yes/no), four features were

extracted—the absence and attendance rates for each attendance type. Figure 3 provides a taxonomy of the student performance factors, subfactors, and the features/attributes used in this study, available from the university dataset.

Table 4. Summary of extracted traditional features.

No.	Traditional Factor	Features/Attributes
1	Demographic Profile	Gender; Religion; Race; Nationality; Local_International and Age
2	Academic History/Educational Background	School type; Qualification achieved; and Number of completed qualifications
3	Scholarship	Having a scholarship (Y/N)
4	Family Background	Whether the student is a first-generation student (Y/N) and parents' occupations
5	Attendance	Lecture attendance and absence rate, tutorial attendance and absence rate
6	Academic Performance	Final mark

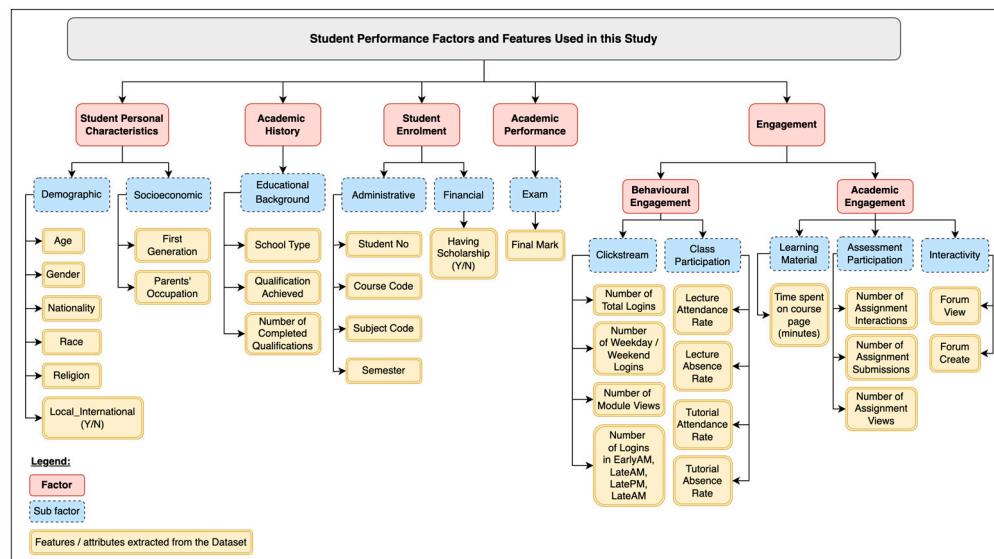


Figure 3. Taxonomy of features used in this study and their corresponding factors.

2.3. Tools and Statistical Test Used

This study utilized the Scikit-learn framework using Python v3.11.4 programming language via the Anaconda IDE application with Jupyter Notebook v6.5.4. The framework includes various libraries used in pre-processing and in data visualization and to perform most of the experiments required to perform statistical analysis.

Exploratory data analysis (EDA) is an approach employed to analyze datasets and summarize their main characteristics, often with visual methods, without making any assumptions about their contents (Rajuladevi, 2018). It is a crucial step to perform before diving into statistical modeling or analysis. EDA was conducted using descriptive analysis and correlation tools such as heatmaps.

Following correlation analysis, non-parametric tests, also known as distribution-free statistical tests, were used when the assumption of normal data distribution could not be made. These tests were used when parametric tests could not be applied (Tilak & Arivazhahan, 2022). However, this study revealed that normality was not an issue due to

the large sample size, where, in most cases, the results of non-parametric and parametric tests were close to one another.

The Mann–Whitney U test, also known as the Wilcoxon rank-sum test, was used for the comparison of two independent groups (Richardson, 2018; Hussain & Mahmud, 2019) such as gender. Kruskal–Wallis H test (Schuster, 1985) was used to compare the distribution of a continuous variable (e.g., “Final Mark”), in this case, across multiple groups such as nationalities.

Following the results of the above tests, post hoc analysis and pairwise tests were used. These are commonly used when the analysis of variance results in a statistically significant difference between pairs of groups. Researchers are mainly interested in identifying the specific groups contributing to the significant value. In this case, Tukey’s Honest Significant Difference (HSD) or Dunn’s tests were more suitable (Nanda et al., 2021; Terpilowski, 2019; Chmiel et al., 2022).

3. Results

This section presents the results and key findings of the analysis and statistical tests to answer the research questions highlighted in Section 1.1 (please refer to Table 5).

Table 5. Research questions aligned with the Sections of this paper.

Research Question (RQ)	Answered in Section
RQ1	Sections 3.1 and 3.2
RQ2	Section 3.3
RQ3	Sections 3.4–3.6
RQ4	Section 4

3.1. The Impact of Demographic and Academic History Factors on Academic Performance

This dataset had 175 male students and 46 female students. Using Tukey HSD, the tests revealed a negative mean difference of approximately -16.9013 units, indicating that, on average, male students had a lower attendance rate than female students. The adjusted p -value was less than 0.05, indicating that the difference in mean attendance rates was statistically significant. This suggests that the difference was unlikely to have occurred by random chance alone. In summary, evidence suggested a statistically significant difference in mean attendance rates between female and male students. The negative mean difference indicated that, on average, female students had a higher attendance rate in tutorials than male students. This was further supported by the Mann–Whitney U test, a non-parametric test that revealed a p -value of 0.0323.

The bar plot in Figure 4 compares the average final mark across different categories, such as male and female. This provides a quick overview of the performance using the gender variable. The findings reveal that, on average, female students scored higher or performed better in exams than male students.

Before exploring these findings in more detail to verify the plot using statistical tests, normality tests needed to be performed to check whether there was a normal distribution between each group (male with final mark) and (female with final mark). For this purpose, the Shapiro–Wilk test was conducted. It was found that since the p -value of the male group was significantly less than the standard significance level of 0.05, it was suggested that this group did not follow a normal distribution. Meanwhile, the female group’s p -value was larger than 0.1705, indicating that it fell within the normal distribution.

Therefore, due to the above findings, non-parametric tests that were robust to violations of normality needed to be performed. In this case, the Mann–Whitney U Test

was performed to measure the significant difference between the distribution of the two independent groups illustrated in Table 6.

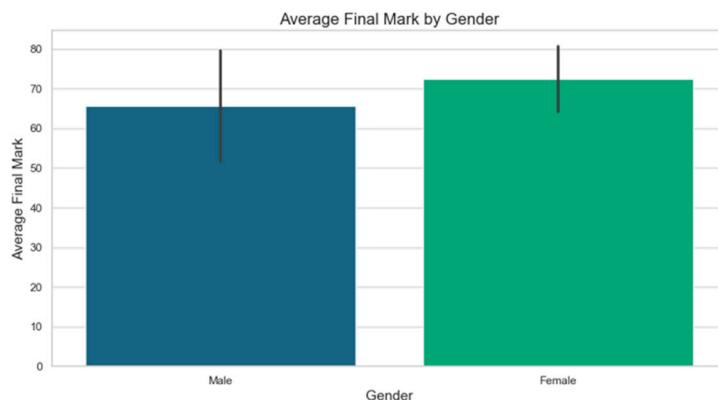


Figure 4. Bar plot of average final marks by gender.

Table 6. Mann–Whitney U test results.

Independent Variable	U-Statistic	p-Value
Gender [Male, Female]	18,607.5	1.3201×10^{-6}
Scholarship [Yes, No]	43,924.5	0.0001

Dependent variable: Final_Mark.

As the *p*-value was less than the significance level (<0.05), we can reject the null hypothesis and conclude that there was a significant difference between the two groups.

The Mann–Whitney U test, also known as the Wilcoxon rank-sum test, is specifically designed to compare two independent groups. Since “Nationality” consists of several groups, more than two, using different statistical tests, such as ANOVA or Kruskal–Wallis, is required. However, the assumptions of normality and homogeneity must be met to perform the ANOVA test. This was tested for Nationality with the final mark, and the results revealed violation in these terms; therefore, using a nonparametric test such as Kruskal–Wallis was more appropriate.

The Kruskal–Wallis test results indicate significant differences in the “Final_Mark” scores among at least some nationalities, as revealed by the *p*-value of less than 0.05, which supports the rejection of the null hypothesis. The H-statistic measures the variability in the data, which can be attributed to the differences between groups.

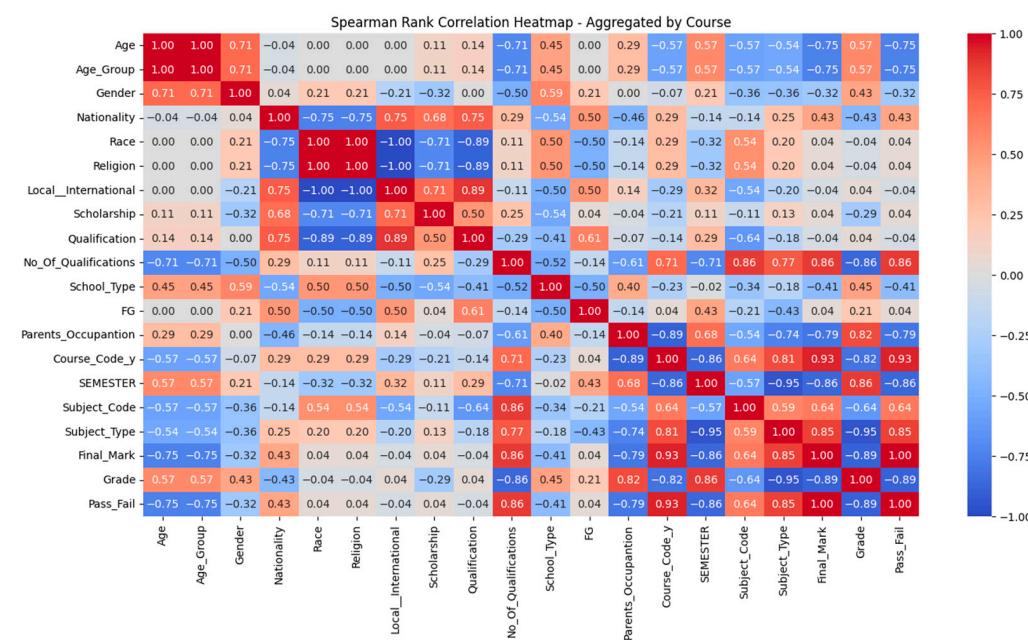
Following these findings, it is interesting to explore which groups are statistically significant; in other words, it is worth finding which nationalities result in statistical differences with the final mark. For this purpose, post hoc tests or pairwise comparisons should be performed to identify which specific groups differ from each other.

Tukey’s Honest Significant Difference (HSD) post hoc test was performed in this case. Table 7 shows the pairwise comparisons between different nationalities for the “Final_Mark” variable. The results included a specific comparison of groups, the mean difference, the *p*-value, and the confidence interval for the mean difference.

Figure 5 represents the correlation heatmap revealing the correlations between all features relating to the demographic, socioeconomic, educational background, administrative, financial, and academic performance factors.

Table 7. Pairwise comparisons of mean differences in final marks among different nationalities.

No.	Group1	Group2	Meandiff	p-adj	Lower	Upper	Reject
1	Libya	Malaysia	21.6952	0.0040	3.6291	39.7613	True
2	Libya	Indonesia	22.9333	0.0119	2.4690	43.3976	True
3	Libya	Maldives	23.2424	0.0312	0.9297	45.5552	True
4	Libya	Norway	-37.8889	0.0150	-72.2574	-3.5203	True
5	Malaysia	Korea	-13.8618	0.0202	-26.7192	-1.0045	True
6	Malaysia	Norway	-39.3618	0.0016	-70.5175	-8.2062	True
7	Indonesia	Norway	-40.6000	0.0021	-73.2049	-7.9951	True
8	Maldives	Norway	-40.9091	0.0035	-74.7048	-7.1134	True
9	India	Norway	-38.8000	0.0268	-75.5832	-2.0168	True

**Figure 5.** Spearman rank correlation heatmap of traditional factors with academic performance.

Using the Spearman test, a correlation coefficient of 0.86 indicates a strong positive linear relationship between the number of qualifications and the final mark. As the number of qualifications increases, the final mark also tends to increase, as illustrated in Figure 5.

Furthermore, correlation analysis on the effect of “age” on academic performance revealed a coefficient between age and final marks of −0.75. This negative correlation suggests a moderately strong inverse relationship between age and final marks. In other words, final marks tend to decrease as age increases, and vice versa. The strength of the correlation indicates a relatively consistent pattern.

3.2. The Impact of Parental Occupation on Students’ Exam Performances

Recognizing the influence of parental occupation and involvement is essential to support a student’s academic success. In this study, we studied the impact of different parents’ occupations on achieving good performance. Figure 6 represents a WordCloud demonstrating the top three occupations of students’ parents: business, housewife, and manager.

Parents’ occupations were assessed based on the student’s academic performance using the Kruskal test, which revealed an H-Statistic of 95.413 and a *p*-Value of 0.023, less than the significance level of 0.05. This suggests a statistical difference between the student’s performance and different groups of parents’ occupations. Furthermore, post hoc analysis was conducted using Dunn’s test, which revealed the following results in Table 8.

Lectures generally have higher attendance rates compared to tutorials. The lecture absence rate has a moderately positive correlation of 0.59 with the tutorial absence rate, which suggests a moderate tendency for students who miss lectures to also miss tutorials.

Before analyzing the effect of attendance on student's performance, normality tests were conducted using the Q-Q (Quantile–Quantile) plot, which is a graphical tool used to assess whether a dataset follows a specific theoretical distribution, such as the normal distribution. It compares the quantiles of the observed data against the quantiles of the expected distribution. As a result of Figure 7 and the Shapiro–Wilk statistical test, the dependent variable, which is the final mark, does not follow a normal distribution. This is due to some points departing from the straight line and the p -value being less than 0.05, which is relatively small.

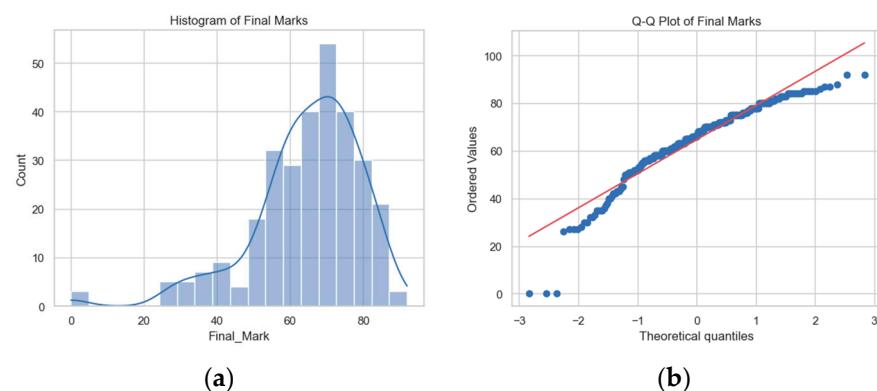


Figure 7. Normality test for final mark: (a) histogram; (b) Q-Q plot.

Due to the above findings, non-parametric tests are more suitable, as they are less sensitive to outliers, work well with small sample sizes, and do not assume a specific distribution for the underlying population. In other words, they can capture a non-linear relationships as well. Instead of relying on the Pearson correlation, the Spearman rank correlation is performed to test the relationship between attendance rates and final marks, as shown in Figure 8.

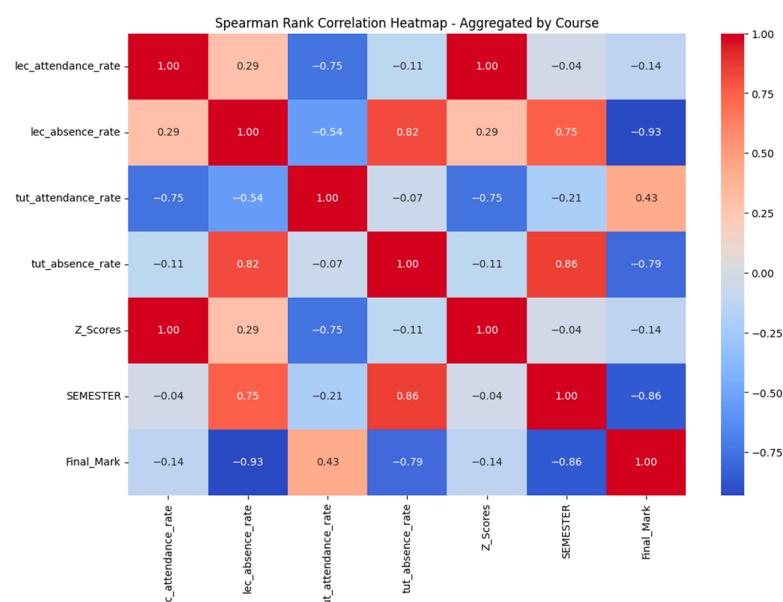


Figure 8. Spearman rank correlation heatmap of attendance with final mark.

In summary, the tutorial attendance rate has a statistically significant positive effect on final marks, while the lecture attendance rate does not show a significant monotonic relationship. Furthermore, the attendance rate for tutorials correlates positively with students' exam final mark of 0.43. This suggests that as the tutorial attendance increases, the final mark increases. In other words, students who attend tutorial classes are more likely to score well in the final exam.

3.4. Assessing LMS Login Behaviour with Attendance and Exam Performance

Figures 9–11 provide an intuitive way to understand the relative proportion of logins during weekdays (coral sector) compared to weekends (blue sector). The pie charts grasp the temporal distribution of average logins and identify patterns in login behavior based on the day of the week.

Proportion of Logins During Different Times of the Day

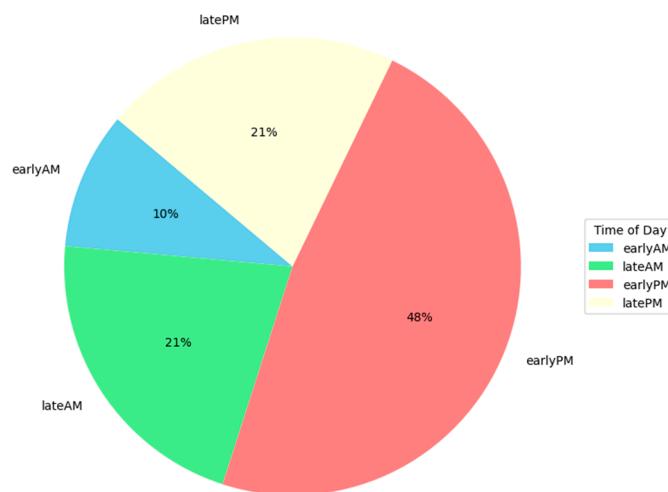


Figure 9. Proportion of logins during different times of the day.

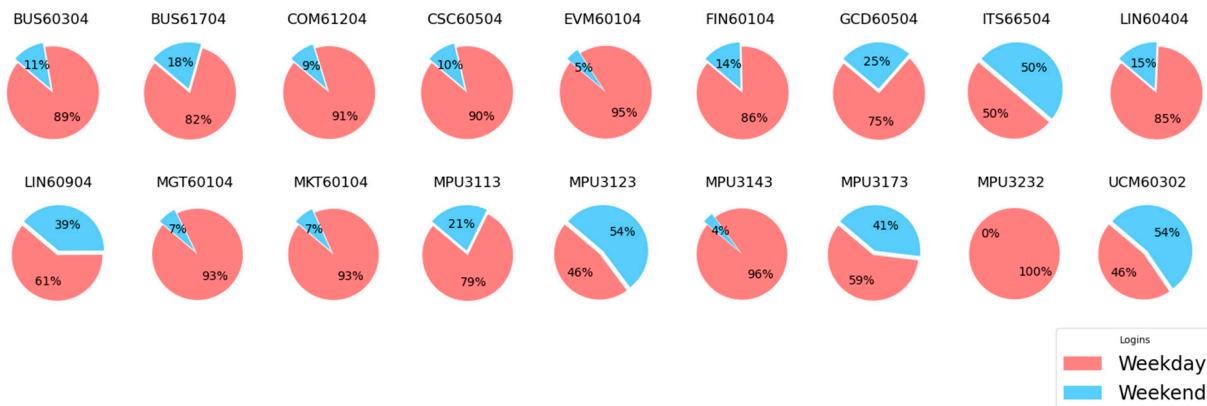


Figure 10. Avg. weekday vs. weekend logins among 19 undergraduate modules.

As illustrated in Figure 12a, tutorial attendance correlates positively with weekend logins by having a strong 0.79 correlation.

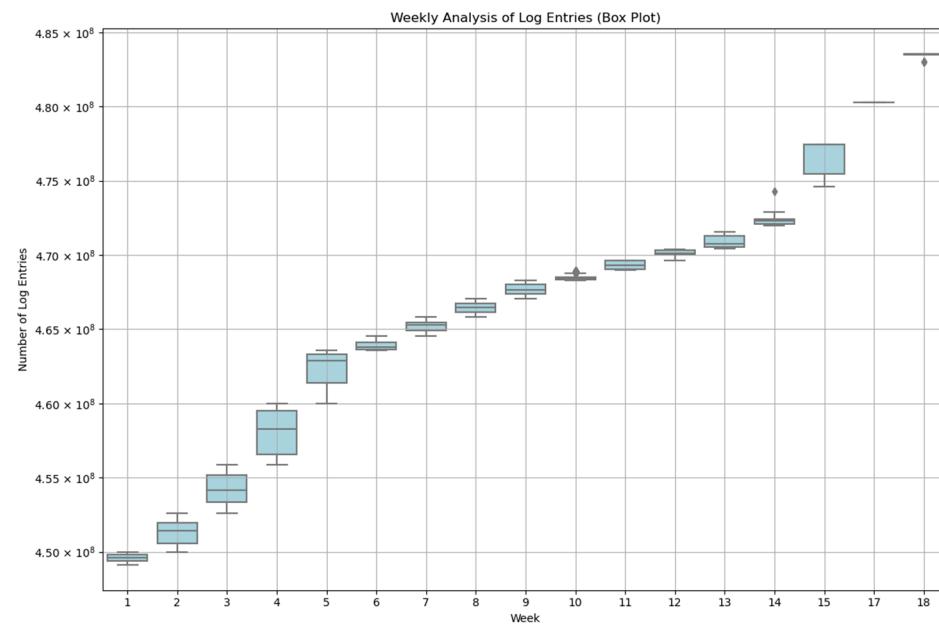


Figure 11. Weekly analysis of log entries.

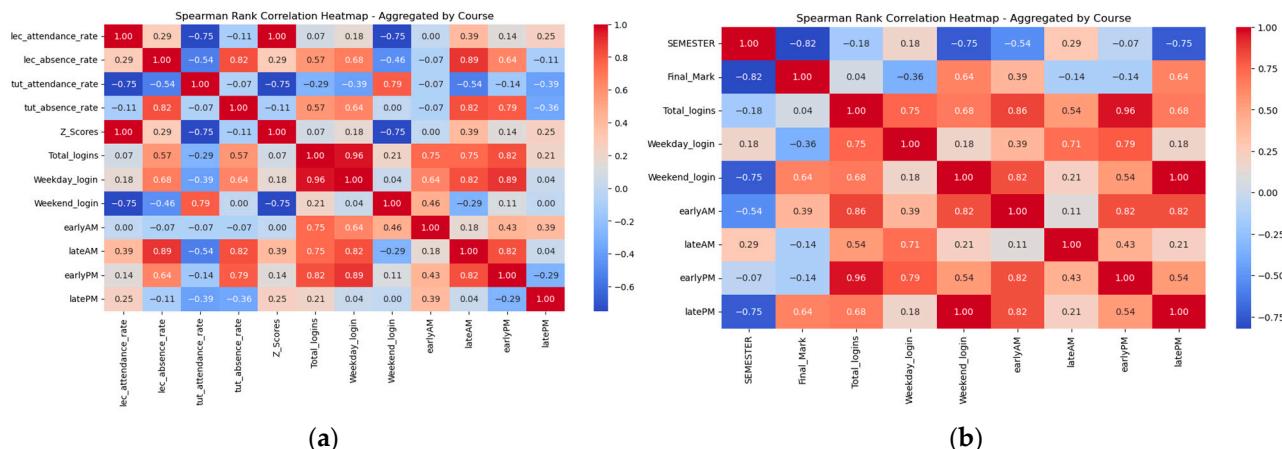


Figure 12. Spearman rank correlation heatmap of (a) LMS engagement (clickstream) with attendance; (b) LMS engagement (clickstream) with final mark.

Furthermore, weekend logins and final exam marks positively correlate, using a Spearman correlation coefficient of 0.64. This positive correlation indicates that students who engage more on weekends (higher weekend logins) tend to achieve higher final exam marks. This could imply a positive relationship between online engagement during weekends and academic success. This is illustrated in Figure 12b.

Regarding hourly login, a Spearman correlation coefficient of 0.64 between late PM logins (19:59 to 23:59) and the final exam mark indicates a moderately strong positive monotonic relationship between these two variables. Students who engage more during late evening hours (late PM logins) tend to achieve higher final exam marks. This could imply a positive relationship between night-time online engagement and academic success.

3.5. Assessing LMS Engagement Behaviour with Attendance and Exam Performance

To investigate the different patterns of LMS logins and engagement among students who passed and failed, a Kruskal–Wallis test was performed, where the passing score was assumed to be set to 50 out of 100. In Figure 13, the bar plot shows the obtained p -values

based on the significance levels (<0.05), which is indicated by the red dashed line for all features extracted from Moodle.

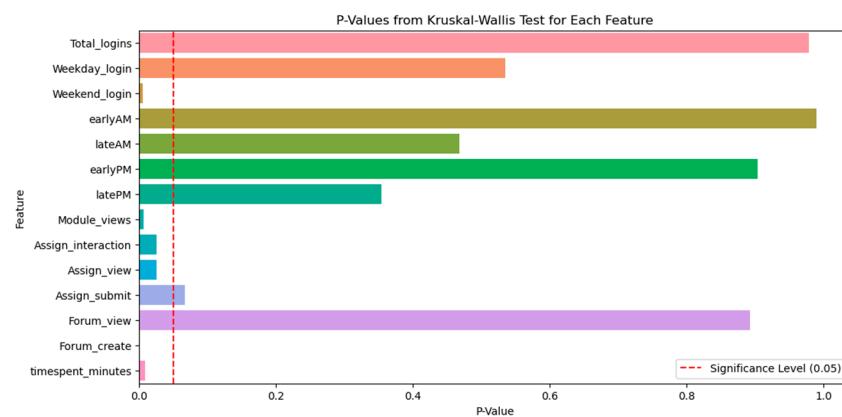


Figure 13. Overall p -Values from Kruskal–Wallis test between pass and fail students.

Moreover, as supported by the Figure, we can conclude that actively engaging with the module on the platform, such as viewing the page, contributes to better understanding and performance. This is indicated by the positive p -value of 0.006.

Students who obtained a passing score in their module may be more engaged with peer interaction and collaborative learning. This is indicated by the 2.48×10^{-8} p -value on the “forum create” feature. This group of students is more likely to create or initiate discussion in the online course forum.

In terms of login behavior, we identified that the total number of logins to the platform and weekday logins does not differ significantly between “Pass” and “Fail” students. Figure 14a,b reveal the correlation of LMS academic engagement with the attendance and final mark, respectively.

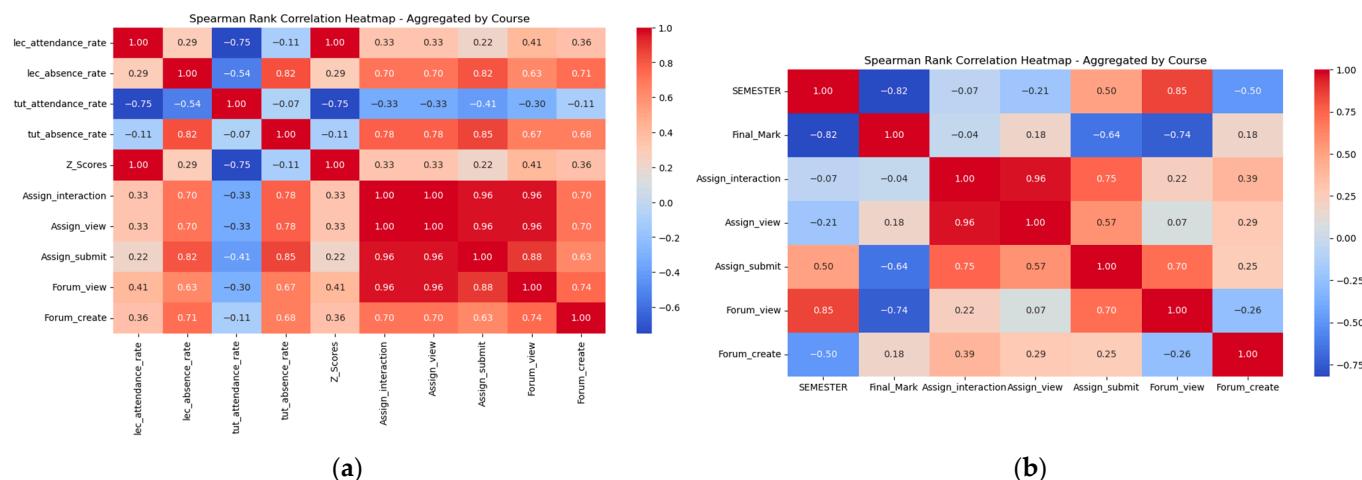


Figure 14. Spearman rank correlation heatmap of: (a) LMS academic engagement with attendance; (b) LMS academic engagement with final mark.

However, weekend logins were statistically significant among the groups, with a p -value of 0.0053. This was further investigated by taking a quantitative approach using box plots to calculate the difference in median final marks among students who logged in during the weekend and those who did not, as demonstrated in Figure 15.

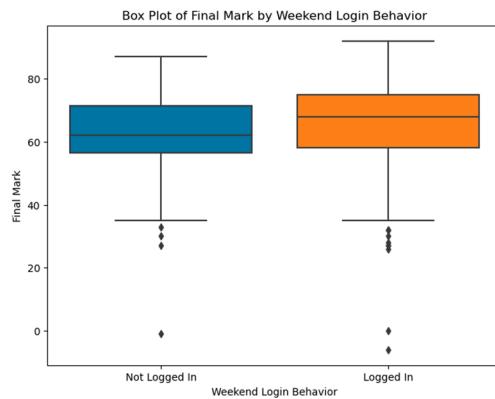


Figure 15. Distribution of final marks by weekend login behavior.

3.6. Assessing Time Spent on the LMS with Attendance and Exam Performance

To study whether the total average time spent daily by students in minutes affected their attendance and exam performance, a heatmap was used to study the correlation, using the Spearman rank test as illustrated in Figure 16a,b.

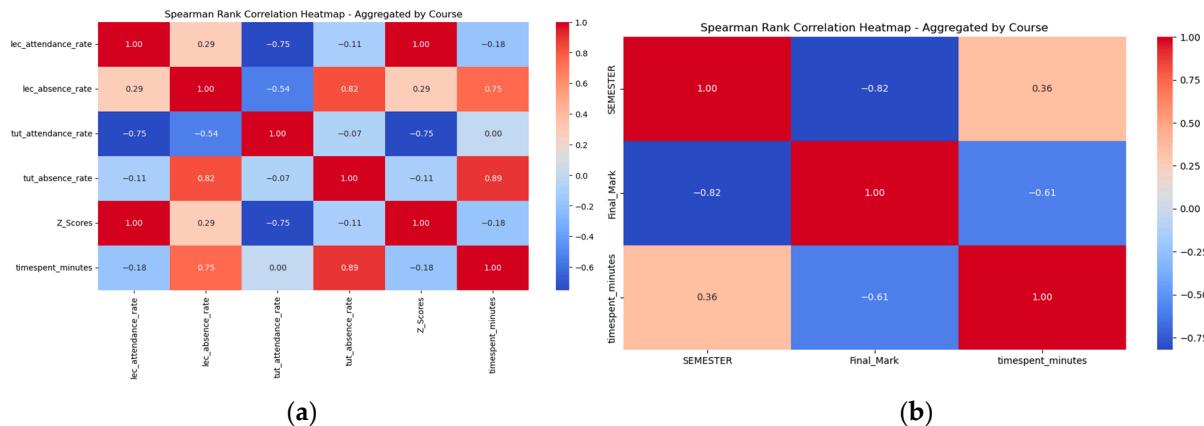


Figure 16. Spearman rank correlation heatmap of (a) time spent on course page with attendance; (b) time spent on course page with final marks.

Lastly, the time spent on the platform in minutes among “pass” and “fail” students showed a statistically significant difference, with a p -value of 0.008. This emphasizes that investing more time in accessing the module materials and engaging in different activities on the platform may lead to student success.

4. Discussion

In this section, we discuss the results and their implications, highlighting the contributions made by this study to the existing literature and extensively addressing the research questions.

The results of our study indicate that students with a higher number of qualifications tend to achieve higher final marks. This positive correlation suggests a valuable finding, as it provides insights into the relationship between academic qualifications and exam performance. If this relationship holds, it might be worthwhile for educational institutions to consider the number of qualifications as a potential factor influencing student success, suggesting that prior academic experience may contribute to overall performance.

Our study also investigated whether cultural differences impact student performance, as revealed by the findings on the Nationality of students. Given that some international students, such as those from Libya, may not be familiar with the Malay language or

its cultural context, this could present additional challenges for them in adapting to the teaching and learning atmosphere and achieving higher marks in subjects that emphasize these aspects, such as MPU modules. This supports the conclusion that there are actual disparities in academic performance among these groups, and the findings are not merely due to chance, as indicated by the *p*-value. Such insights can help educational institutions better support international students.

Attendance count and whether the student takes a leave of absence are strong predictors of student performance. In this study, it was revealed that, overall, female students had better class attendance than male students; this was in line with [Mitra et al.'s \(2022\)](#) results, where a positive correlation was revealed. However, our findings offer new insights into analyzing student performance according to different class types rather than focusing on overall attendance. For example, this study investigated how traditional factors incorporated with online LMS features are used to understand student engagement in a bigger picture. We reveal that more weekend logins indicate higher tutorial attendance, which could have many implications related to engagement in learning. The relationship is said to be monotonic, meaning an increase in both variables; however, it is not necessarily at a constant rate. This might be due to the inconsistent trends in students' logins over time.

In this study, it was found that students attending lectures are more likely to participate in tutorials. However, other factors might contribute to this correlation, and further analyses or experiments would be needed to establish causation.

The above findings suggest that students who attend tutorials regularly are more likely to be engaged in the learning process and demonstrate higher levels of weekend login activity, emphasizing their commitment to academic success, as they are proactive about their studies during weekends, possibly working on assignments, reviewing materials, or seeking additional resources online. The positive correlation highlights a consistent pattern of learning behavior among students. Those who invest time in tutorials during the week may carry this commitment into the weekend, maintaining a steady approach to their academic responsibilities.

Furthermore, our study revealed that tutorial attendance is more prominent in impacting exam performance. This highlights the importance of a holistic approach to student engagement, where it suggests that students who passed their module exhibit a higher frequency of logins outside of their class hour, as they tend to put more effort into reviewing the module materials and completing their tutorials during the weekend. Further investigation revealed that, on average, students who regularly log on to the platform on the weekends tend to achieve a final mark of 6.0 points higher than those who do not, representing a 10% relative increase in their score.

In research, it has been well documented that parental occupation has a significant impact on students' academic performance. Prior studies reveal that students whose parents work in professional fields tend to have higher academic achievements ([Hussain & Obaid, 2022](#); [Roshita et al., 2023](#)), a trend mirrored in the findings of our study as well. More specifically, it highlighted the superior performance of students whose parents work in "Business" roles, which may be associated with high educational expectations, instilling values of diligence and perseverance in their upbringings. Conversely, other occupations such as "Air traffic control", "Employee", "District officer", or "Executive" tend to affect the performance of students, where they might face academic challenges. These results emphasize the critical influence of parents' occupations on educational outcomes.

Furthermore, age has been considered by many studies as a feature that can be used to analyze and predict academic performance. However, our study reveals an inverse correlation between age and the final mark, with older individuals tending to have lower performance. This was further supported by the findings of [Bravo-Agapito et al.'s \(2021\)](#)

study, indicating age as a negative predictor of performance. This may be due to the decline in cognitive mechanisms necessary for learning something new that occurs with age (Ameen et al., 2019). Therefore, while age may have an impact on academic performance, it is not the sole determining factor, and other individual characteristics that were beyond the scope of this study should also be considered.

Traditional factors such as student's demographics and socioeconomic background may have some influence on rationalizing differential outcomes and identifying the performance of students, but it must be used with caution (You, 2016). However, this does not eliminate the necessity for the inclusion of these features in prediction modeling to facilitate robust profiling studies of academic performance. In terms of demographic features such as gender, our study indicates that female students tend to perform better in final exams, and having a scholarship may be associated with a statistically significant impact on final marks in the dataset, suggesting that students with scholarships tend to perform better in final exams.

Forum tools on online educational platforms are utilized to anticipate participation and discussion among students. This feature has remained persistent in the existing literature, as it can strongly predict a student's performance (Shafiq et al., 2023). Active participation in forums fosters knowledge sharing, problem solving, and community building, positively impacting learning outcomes. Our study contributes to the literature by demonstrating that creating discussion in forums is significant among pass and fail students, suggesting that educational institutions should put more emphasis on this tool to improve learning practices. This was in line with the results found by Pa et al. (2023) on the factors related to predicting the final results, which suggested that more focus on forum discussion is required.

Overall, these findings highlight the importance of active engagement, participation, and time investment in academic success. Students who pass tend to demonstrate higher levels of interaction with course materials, assignments, and peers, suggesting a more comprehensive and dedicated approach to their studies. By contrast, students who fail may benefit from strategies to increase their engagement and involvement with course content, assignments, and collaborative learning opportunities to improve their chances of success. This was in line with Jovanović et al.'s (2021) findings, where it was concluded that the total time spent online, actively contributing to the forum, and access to course materials proved to be the most significant predictors of students' learning outcomes (measured through the final course grade) across courses.

Among the login behavior features, it is concluded that the total number of logins to the platform and weekday logins do not affect academic performance. This could be explained by the instructional design of the courses in the university, where students are required to access the platform on a weekly basis, often during their classes, whether low- or high-performing students; therefore, the results did not reveal any significance for such features. However, total logins to the platform may be useful in understanding student online engagement patterns; for example, our results revealed that modules such as MPU and UCM that are heavy on tutorial tasks may exhibit higher course logins by students on weekends. Furthermore, we notice that the number of log entries tends to increase during week 15, assuming the modules in this study expand over a 14-week period for one semester in the university; week 15 is where students may likely refer more to course materials to study in an attempt to prepare for their exams.

Furthermore, as revealed by the results, there is a positive correlation between the class absence rate and the time spent on the e-learning platform, which may seem counter-intuitive at first glance; however, it could be due to several factors such as compensatory behavior, where students who are often absent for class may feel the need to compen-

sate and spend more time to catch up on the missed materials, or it could be due to the complexity of the course material.

Ultimately, it is found that module/course views, forum creation, overall assignment interaction, and time spent on the platform are among the top LMS variables that contribute significantly to students' academic performance. The analyzed features seem to complement each other, such as tutorial attendance and login activity on the weekend. While the importance of such variables has already been widely mentioned in the literature (Li et al., 2022; Arizmendi et al., 2023; Alnasyan et al., 2024; Wang & Mousavi, 2023), the latter features related to the weekend and hourly logins received less attention (Quinn & Gray, 2019), and their significance across multiple courses is a new contribution that this study brings to the field.

5. Conclusions

The study examined a comprehensive set of online and traditional features covering several aspects of students' learning behavior in an LMS, involving indicators of engagement levels and interaction regularity. Besides the overall analysis of students' online behavior, temporal analysis across multiple courses was performed.

By analyzing log variables alongside traditional factors, future researchers can gain deeper insights into how students engage with e-learning platforms, which in turn helps educators tailor their teaching methods and optimize their educational resources. For example, in this study, it is revealed that instructors need to encourage their students to engage more with the content available on Moodle, especially using the Forum tool, as it can spark discussion and motivate them to score higher in their final exam.

The study is also beneficial to future research, as it uncovers the relationship between online and offline features that could be leveraged to build accurate predictive models in the field of Learning Analytics. Moreover, as the results revealed significant and insignificant features in Moodle logs, such insights can greatly aid future researchers in building effective feature selection methods.

5.1. Limitations

This paper is subject to certain limitations that can be addressed by future research. The study was based on a sample of 19 blended courses that were assumed to be homogeneous regarding institutional settings, discipline, and nominal learning design. Therefore, this is a current limitation of the study, as less information on the courses was provided. Furthermore, several other forms of content on Moodle exist, such as quiz components, videos, and lecture recordings, which were not considered in this study due to their lack of use in the courses involved in this study.

5.2. Future Direction

As e-learning tools evolve, students may exhibit different interaction and behavioral patterns on learning platforms; therefore, in the future, it is worth identifying heterogeneous groups of students to build predictive models with the ability to provide tailored and group-focused interventions. Furthermore, researchers could investigate more patterns beyond the accessibility features in Moodle logs, as this could provide more valuable information on what kind of resources students prefer and make more informed decisions on their teaching and learning. Future research could also investigate the underlying factors contributing to the inverse relationship between age and academic performance, such as external commitments like work and family responsibilities, cognitive changes, or disparities in educational experiences.

Author Contributions: Conceptualization, D.A.S. and M.M.; Methodology, D.A.S.; Software, D.A.S.; Validation, D.A.S., M.M. and R.A.A.H.; Formal analysis, D.A.S.; Investigation, D.A.S.; Resources, D.A.S., M.M. and D.A.; Data curation, D.A.S. and M.M.; Writing—original draft, D.A.S.; Writing—review & editing, D.A.S., M.M., R.A.A.H. and D.A.; Visualization, D.A.S.; Supervision, M.M., R.A.A.H. and D.A.; Project administration, M.M.; Funding acquisition, D.A. All authors have read and agreed to the published version of the manuscript.

Funding: The Article Processing Charge was funded by Taylor's University under grant number IVERSON/2018/SOCIT/001.

Institutional Review Board Statement: Ethical review and approval were waived for this study, as the research did not involve collecting or using personal data that could reveal the identity of any individual participants.

Informed Consent Statement: Participant consent was waived, as the research did not involve collecting or using personal data that could reveal the identity of any individual participants.

Data Availability Statement: The data presented in this study are available on request from the corresponding author (The dataset used in this research belongs to a private university in Malaysia. Data cannot be made public due to restrictions of the authorized university to preserve anonymity and confidentiality).

Acknowledgments: The authors would like to thank Taylor's University for the opportunity to carry out this research under Taylor's Research Excellence Scholarship 2021.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Abu Zohair, L. M. (2019). Prediction of student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, 16(1), 27. [\[CrossRef\]](#)
- Aguagallo, L., Salazar-Fierro, F., García-Santillán, J., Posso-Yépez, M., Landeta-López, P., & García-Santillán, I. (2023). Analysis of student performance applying data mining techniques in a virtual learning environment. *International Journal of Emerging Technologies in Learning (IJET)*, 18(11), 175–195. [\[CrossRef\]](#)
- Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, 31(1), 542–550. [\[CrossRef\]](#)
- Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, 16(1), 40. [\[CrossRef\]](#)
- Alnasyan, B., Basher, M., & Allassafi, M. (2024). Deep learning techniques for predicting student's academic performance on virtual learning environments: A review. *Research Square*. [\[CrossRef\]](#)
- Al-Sulami, A., Al-Masre, M., & Al-Malki, N. (2023). Predicting at-risk students' performance based on lms activity using deep learning. *International Journal of Advanced Computer Science and Applications*, 14(6). [\[CrossRef\]](#)
- Ameen, A. O., Alarape, M. A., & Adewole, K. S. (2019). Students' academic performance and dropout prediction. *Malaysian Journal of Computing*, 4(2), 278. [\[CrossRef\]](#)
- Arizmendi, C. J., Bernacki, M. L., Raković, M., Plumley, R. D., Urban, C. J., Panter, A. T., Greene, J. A., & Gates, K. M. (2023). Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work. *Behavior Research Methods*, 55(6), 3026–3054. [\[CrossRef\]](#)
- Bainbridge, J., Melitski, J., Zahradník, A., Lauría, E. J. M., Jayaprakash, S., & Baron, J. (2018). Using learning analytics to predict at-risk students in online graduate public affairs and administration education. *Journal of Public Affairs Education*, 21(2), 247–262. [\[CrossRef\]](#)
- Banda, L. O. L., Liu, J., Banda, J. T., & Zhou, W. (2023). Impact of ethnic identity and geographical home location on student academic performance. *Helyon*, 9(e16767), 1–18. [\[CrossRef\]](#)
- Bravo-Agapito, J., Romero, S. J., & Pamplona, S. (2021). Early prediction of undergraduate student's academic performance in completely online learning: A five-year study. *Computers in Human Behavior*, 115, 106595. [\[CrossRef\]](#)
- Chmiel, D., Wallan, S., & Haberland, M. (2022). tukey_hsd: An accurate implementation of the Tukey honestly significant difference test in Python. *Journal of Open Source Software*, 7(78), 4383. [\[CrossRef\]](#)
- Cui, Y., Chen, F., & Shiri, A. (2020). Scale up predictive models for early detection of at-risk students: A feasibility study. *Information and Learning Sciences*, 121(3/4), 97–116. [\[CrossRef\]](#)

- Çakiroğlu, Ü., Kokoç, M., & Atabay, M. (2024). Online learners' self-regulated learning skills regarding LMS interactions: A profiling study. *Journal of Computing in Higher Education*, 36(4), 220–241. [CrossRef]
- Fernández-Leyva, C., Tomé-Fernández, M., & Ortiz-Marcos, J. M. (2021). Nationality as an influential variable with regard to the social skills and academic success of immigrant students. *Education Sciences*, 11(10), 605. [CrossRef]
- Furqon, M., Sinaga, P., Liliasari, L., & Riza, L. S. (2023). The impact of Learning Management System (LMS) usage on students. *TEM Journal*, 12(2), 1082–1089. [CrossRef]
- Gamage, S. H. P. W., Ayres, J. R., & Behrend, M. B. (2022). A systematic review on trends in using Moodle for teaching and learning. *International Journal of STEM Education*, 9(1), 9. [CrossRef]
- Gonzalez-Nucamendi, A., Noguez, J., Neri, L., Robledo-Rella, V., García-Castelán, R. M. G., & Escobar-Castillejos, D. (2022). Learning analytics to determine profile dimensions of students associated with their academic performance. *Applied Sciences*, 12(20), 10560. [CrossRef]
- Halal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J., & Long, Q. (2018). Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems* 161, 134–146. [CrossRef]
- Hussain, M., & Mahmud, I. (2019). pyMannKendall: A Python package for non-parametric Mann-Kendall family of trend tests. *Journal of Open Source Software*, 4(42), 1556. [CrossRef]
- Hussain, M., & Obaid. (2022). The role of mother's profession on academic performance of university students. *Pakistan Journal of Social Research*, 4(1), 140–149. [CrossRef]
- Ifeanyi, U. N., & Chinonso, S. (2023). The impact of learning management system on student academic performance of computer science department of federal polytechnic Kaura Namoda, Zamfara state, Nigeria. *Asian Journal of Research in Computer Science*, 16(3), 11–17. [CrossRef]
- İşikgöz, M. E. (2024). Do learning management system activities in online pedagogical education significantly predict academic performance? *Turkish Online Journal of Educational Technology*, 23(1), 53–60.
- Jiang, Y. (2022, November 26–28). *Statistical analysis of several factors in predicting student performance*. Proceedings of the International Conference on Statistics, Applied Mathematics, and Computing Science (CSAMCS 2021), Nanjing, China.
- Jovanović, J., Saqr, M., Joksimović, S., & Gašević, D. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computers & Education*, 172, 104251. [CrossRef]
- Khairy, D., Alharbi, N., Amasha, M. A., Areed, M. F., Alkhalfaf, S., & Abougalala, R. A. (2024). Prediction of student exam performance using data mining classification algorithms. *Education and Information Technologies*, 29(16), 21621–21645. [CrossRef]
- Khan, Y. L., Lodhi, S. K., Bhatti, S., & Ali, W. (2019). Does absenteeism affect academic performance among undergraduate medical students? Evidence from Rashid Latif Medical College (RLMC). *Advances in Medical Education and Practice*, 10, 999–1008. [CrossRef]
- Khor, E. T., & Dave, D. (2022). A Learning Analytics Approach Using Social Network Analysis and Binary Classifiers on Virtual Resource Interactions for Learner Performance Prediction. *The International Review of Research in Open and Distributed Learning*, 23(4), 123–146. [CrossRef]
- Kondo, N., Okubo, M., & Hatanaka, T. (2017, July 9–13). *Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data*. 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI) (pp. 198–201), Hamamatsu, Japan. [CrossRef]
- Lavidas, K., Komis, V., & Achriani, A. (2022). Explaining faculty members' behavioral intention to use learning management systems. *Journal of Computers in Education*, 9(4), 707–725. [CrossRef]
- Li, C., Herbert, N., Yeom, S., & Montgomery, J. (2022). Retention factors in stem education identified using learning analytics: A systematic review. *Education Sciences*, 12(11), 781. [CrossRef]
- Llerena-Izquierdo, J., Rodriguez, M. E., & Guerrero-Roldán, A.-E. (2023, July 26–28). *Monitoring and adaptation of assessment activities in a VLE supported by learning analytic*. CITIS: International Conference on Science, Technology and Innovation for Society (pp. 409–419), Guayaquil, Ecuador. [CrossRef]
- Loan, D. T. T., Tho, N. D., Nghia, N. H., Chien, V. D., & Tuan, T. A. (2024). Analyzing students' performance using fuzzy logic and hierarchical linear regression. *International Journal of Modern Education and Computer Science*, 16(1), 1–10. [CrossRef]
- Mitra, S., Sarkar, P., Bhattacharyya, S., & Basu, R. (2022). Absenteeism among undergraduate medical students and its impact on academic performance: A record-based study. *Journal of Education and Health Promotion*, 11, 414. [CrossRef]
- Moodle. (2024). *About moodle*. Available online: https://docs.moodle.org/403/en/About_Moodle (accessed on 1 April 2024).
- Mubarak, A. A., Cao, H., & Hezam, I. M. (2021). Deep analytic model for student dropout prediction in massive open online courses. *Computers and Electrical Engineering*, 93, 107271. [CrossRef]
- Mwalumbwe, I., & Mtebe, J. S. (2017). Using learning analytics to predict students' performance in moodle learning management system: A case of mbeya university of science and technology. *The Electronic Journal of Information Systems in Developing Countries*, 79(1), 1–13. [CrossRef]
- Nanda, A., Mohapatra, B. B., & Mahapatra, A. P. K. (2021). Multiple comparison test by Tukey's honestly significant difference (HSD): Do the confidence level control type I error? *International Journal of Statistics and Applied Mathematics*, 6(1), 59–65. [CrossRef]

- Ng, H., bin Mohd Azha, A. A., Yap, T. T. V., & Goh, V. T. (2022). A machine learning approach to predictive modelling of student performance. *F1000Research*, 10, 1144. [[CrossRef](#)]
- Pa, N. N. N., Aziz, A. A., Safei, S., & Mustafa, W. A. (2023). Predicting student final examination result using regression model based on student activities in learning management system. *Journal of Autonomous Intelligence*, 7(1), 1–14. [[CrossRef](#)]
- Prakash, A. (2023, March 25–29). [DC] Investigating student's learning processes in a Virtual Reality learning environment by analyzing their interaction behavior. 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 995–996), Shanghai, China. [[CrossRef](#)]
- Quinn, R. J., & Gray, G. (2019). Prediction of student academic performance using Moodle data from a Further Education setting. *Irish Journal of Technology Enhanced Learning*, 5(1). [[CrossRef](#)]
- Rajuladevi, A. (2018). *A machine learning approach to predict first-year student retention rates at university of Nevada, Las Vegas* [Master's Thesis, University of Nevada, Las Vegas]. UNLV Theses, Dissertations, Professional Papers, and Capstones. Available online: <https://digitalscholarship.unlv.edu/thesesdissertations/3315/> (accessed on 1 April 2024).
- Richardson, J. T. E. (2018). Kruskal–Wallis test. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation*. SAGE Publications. [[CrossRef](#)]
- Roshita, P., Achwan, R., Setianingsih, R., & Dari, P. W. (2023). Investigation of the impact of parents' occupation on the academic grades of high school students. *International Journal of Education Teaching Zone*, 2(2), 264–274. [[CrossRef](#)]
- Schuster, E. (1985). A comparison of algorithms for the computation of the Mann-Whitney test. *Biometrical Journal*, 27(4), 405–410. [[CrossRef](#)]
- Shafiq, D. A., Marjani, M., Habeeb, R. A. A., & Asirvatham, D. (2022). Student retention using educational data mining and predictive analytics: A systematic literature review. *IEEE Access*, 10(2), 72480–72503. [[CrossRef](#)]
- Shafiq, D. A., Marjani, M., Habeeb, R. A. A., & Asirvatham, D. (2023, October 14–15). *A narrative review of students' performance factors for learning analytics models*. Proceedings of the 7th International Joint Conference on Advances in Computational Intelligence (IJCACI 2023) (pp. 273–284), New Delhi, India.
- Shaimov, N. D., Lomazova, I. A., Mitsyuk, A. A., & Samonenko, I. Y. (2022). Analysis of students' academic performance using LMS event logs. *Modeling and Analysis of Information Systems*, 29(4), 286–314. [[CrossRef](#)]
- Shou, Z., Xie, M., Mo, J., & Zhang, H. (2024). Predicting student performance in online learning: A multidimensional time-series data analysis approach. *Applied Sciences*, 14(6), 2522. [[CrossRef](#)]
- Sommerville, T., & Singaram, V. S. (2018). Exploring demographic influences on students' academic performance over a five-year programme. *South African Journal of Higher Education*, 32(2), 273–287. [[CrossRef](#)]
- Suay, A. P., Van Vaerenbergh, S., Piles, M., Laparra, V., Pascual-Venteo, A. B., Ruescas, A. B., Fernandez-Moran, R., Martinez-Garcia, M., Adsuara, J. E., Amoros, J., Munoz-Mari, J., Epifanio, I., & Fernandez-Torres, M.-A. (2022, September 29–30). *Learning about student performance from moodle logs in a higher education context*. 2022 XII International Conference on Virtual Campus (JICV) (pp. 1–4), Arequipa, Peru. [[CrossRef](#)]
- Terpilowski, M. A. (2019). scikit-posthocs: Pairwise multiple comparison tests in Python. *Journal of Open Source Software*, 4(39), 1556. [[CrossRef](#)]
- Tilak, A., & Arivazhahan, A. (2022). Non-parametric tests. In M. Lakshmanan, D. G. Shewade, & G. M. Raj (Eds.), *Introduction to basics of pharmacology and toxicology*. Springer. [[CrossRef](#)]
- Viswanathan, S., & Kumar, S. V. (2021). Study of students' performance prediction models using machine learning. *Turkish Journal of Computer and Mathematics Education*, 12(2), 3085–3091.
- Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*, 104(11), 106189. [[CrossRef](#)]
- Wang, Q., & Mousavi, A. (2023). Which log variables significantly predict academic achievement? A systematic review and meta-analysis. *British Journal of Educational Technology*, 54(1), 142–191. [[CrossRef](#)]
- Wong, B. T., & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education*, 7(1), 7–28. [[CrossRef](#)]
- You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet and Higher Education*, 29(3), 23–30. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.