

Student Performance Prediction Using Virtual Learning Environment (VLE) Interactions

M.S. Faathima Fayaza^{1, 2} and Supunmali Ahangama¹

¹ University of Moratuwa, Moratuwa, Sri Lanka

² South Eastern University of Sri Lanka, Sri Lanka

fayazamsf.22@uom.lk

supunmali@uom.lk

Abstract: Technological advancements have reshaped the education landscape through the introduction of digital learning platforms. Although higher education institutions are striving to increase the learning outcome and reduce the dropout rates, they still face challenges. Virtual Learning Environments (VLEs) have become essential platforms for delivering instructional content and assessing student engagement. This study aims to predict the students' learning outcomes for the database management subject using VLE log data. This study utilised 78,175 VLE click events generated by two hundred and forty-seven (247) students in a distance learning environment from a state university in Sri Lanka. The study utilised seven behavioural features, number of unique components, average hour, standard deviation of the hour, average number of days, number of weekend interactions, number of session count, peak study hour and thirty-four learning activity features to predict the learning outcome. From the Exploratory Factor Analysis (EFA) session count, the number of weekend interactions, and the unique components are selected as the most influential behavioural features, grade user report viewed, discussion created, discussion viewed, course viewed, a file has been uploaded, feedback viewed and course module viewed have been selected as the most influential learning activity features. The study utilises traditional Machine Learning approaches such as Random Forest Regressor, Support Vector Machines (SVM), and Deep Learning approaches, Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) to perform the prediction. Among this approaches the Long Short-Term Memory (LSTM) model, a type of RNN outperform other approaches in terms of accuracy, Mean Absolute Error (MAE), and F1 score. The LSTM model achieved 97% accuracy.

Keywords: Deep Learning, Learning Activity, Log Data Analysis, Machine Learning, Performance Prediction, Virtual Learning Environment (VLE)

1. Introduction

Technology advancements introduced during the COVID-19 pandemic introduced a new norm in education. Traditional pedagogical systems are being replaced by the increasing integration of digital platforms like learning management systems (LMSs). Educational institutions have transformed the traditional learning environment into modern digital environments (Rizwan et al., 2025). According to the global statistical report, 77% of higher educational institutions have incorporated digital platforms to teach their core curriculum (Maaliw, 2021). However, low learning outcomes (Rogers et al., 2025), less engagement (Jawad et al., 2022) remain as core challenges for educational institutions using LMSs.

Though the Virtual Learning Environment (VLE) and the LMS have the same features, they however differ from the way they are used (Pinner, n.d.). VLE serves as the basis of most of the e-learning platforms and makes positive impacts on learners and teachers alike (Maaliw, 2021). VLE changes the teaching and learning environment from a physical classroom to online (Ryani Kusumawati, 2024). VLE supports content sharing, learning activities and assessment (Ryani Kusumawati, 2024). Even though e-learning platforms follow the "one size fits all" style, it has failed to understand the individual needs (Maaliw, 2021). However, learners have different learning needs and learning behaviours, and these impact learning outcomes. Further, lower engagement (Jawad et al., 2022), abnormal participation and different adaptability among the learners reduce their learning outcomes (Ryani Kusumawati, 2024). In addition, poor curriculum design and minimal interactivity create more challenges (Ryani Kusumawati, 2024). Digital divide and digital literacy also influence the learning behaviours of the learners (Zakir et al., 2025). These challenges create the necessity to understand the influencing factors in learning outcomes and the early prediction of learning outcomes.

VLE contains a huge volume of student interaction data with the students' details. Machine Learning (ML) and Deep Learning (DL) are subsets of Artificial Intelligence (AI) and have emerged as powerful tools to predict hidden patterns and make predictions. Understanding how students interact with VLEs can provide valuable insights into their learning behaviours and performance. The available literature mainly relies on statistical analysis (Jo et al., 2018)) and applying ML and DL may provide more details of learner's performance (Borna et al., 2024). Unlike the traditional statistical models, ML and DL can also learn from data-making predictions.

In the literature, researchers commonly utilised course design data and LMS data to predict the performance (Liu et al., 2023). The frequently used data include demographic, academic background and learning behaviour to predict the performance (Liu et al., 2023). The challenges in the reusing model build using course related data are difficult to be applied to another course since the course data is heavily dependent on the course design (Liu et al., 2023). However, the clickstream data contains the details of student activity and navigation during the course (Liu et al., 2023). This study used clickstream data to predict the learners' performance. This study utilises VLE log data collected from a module followed by distance learning students from a state university in Sri Lanka. The study utilises the Random Forest Regressor, Support Vector Machines (SVM), Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) to predict the learning outcome. Therefore, this study applies ML and DL techniques to analyse VLE log data to predict learners' learning outcome, enabling timely interventions and personalised learning strategies.

UNESCO's Sustainable Development Goal 4, namely "...ensure inclusive and equitable quality education and promote lifelong learning opportunities for all," intends to provide quality education for all. This study aligns with the Sustainable Development Goal 4. This study aims to improve the education quality by analysing the learners' learning behaviours in the digital environment along with the learning outcome to improve the learners' learning outcome.

2. Literature Review

The Online Learning Platform is capable of storing vast amounts of students' learning data (Borna et al., 2024). VLEs store students' interaction data (Borna et al., 2024), learning resources and activities. The VLE log data contains students' click interactions with the VLE. These details offer the opportunity to predict student performance. However, the difficulties in accurately identifying at-risk and high-achieving students remain a challenge (Borna et al., 2024). VLE logs contain details about students' learning behaviour. However, current literature has a very limited understanding of how these log events represent the learners' learning behaviours (Fazil et al., 2024). By understanding the students' learning patterns, the teachers can support them with more personalised content.

Learners' behaviours in online learning platforms differ from traditional learning environments, where motivation plays a key role in performance (Borna et al., 2024). However, on average, the learners' engagement contributes to predicting learning outcomes (Al-Tameemi et al., n.d.) and engagement plays a significant role compared to personal information. Student behavioural measurements can be divided into device-based or activity based (Borna et al., 2024).

The available studies in the literature used different approaches and different features to predict the learning outcome. They predict and analyse student performance based on different features. These features have been divided into the following groups (Rizwan et al., 2025): first, academic performance features (previous and current education background features such as courses studied, assignments, quizzes, course grade, final grade, course details, exam scores and GPA etc.). Second, demographic features (student's personal biodata information, which consists of the student's details, family details and social data, such as gender, age, job detail, number of family members, study hours, number of friends etc.). Third, behavioural or clickstream features (login time, online spending time, submitted assessments, web page clicking and web page visiting, discussion forums and video interactions, etc.). Fourth, facial and emotional features (facial expressions, head movements, head poses, eye contact and recognition etc.). Fifth, learning activity features respectively (interaction in VLE, learning behavioural and LMS activity data based on click frequency information, etc.).

In the literature, several features are used to predict the learning outcomes. They are, namely, demographic features, assessment grading, number of clicks and final marks (Al-Tameemi et al., n.d.). Researchers used several VLE features discussion forum (Rogers et al., 2025), course material, HTML material, Home page, quiz (Rogers et al., 2025), for their final result to predict the learning outcome (Al-Tameemi et al., n.d.). VLE resources, files, links and forum discussions have positive impacts on learning outcomes (Rogers et al., 2025). The most influential learning behaviour is submission, and the least influential feature is delete in performance prediction among the following features: submission, quiz, forum create, read, update, and delete days logged (Rogers et al., 2025). Course module activity design, collaboration, student engagement with VLE, features use, and adaptability features positively impacted learning outcomes (Ryani Kusumawati, 2024). Among these features, the most influential and the second influential features are the course module activity design.

Learning analytics is a relatively young research field (Alasalmi, 2021). Educational Data Mining (EDM), ML, and DL help to understand the learners' behaviours (Al-Tameemi et al., n.d.). In the literature, the widely used

techniques in the performance predictions are Random Forest (RF) (Rogers et al., 2025; Yağcı, 2022a), Nearest Neighbour (Yağcı, 2022a), Support Vector Machine (SVM) (Yağcı, 2022), Logistic Regression (Altabrawee et al., 2019; Yağcı, 2022b), XGBoost (Rogers et al., 2025), Decision Tree (DT) (Altabrawee et al., 2019; Rogers et al., 2025), Artificial Neural Networks (ANN) (Altabrawee et al., 2019; Ryani Kusumawati, 2024b) and K-Nearest Neighbour (Yağcı, 2022a).

Random Forest, Decision Tree, and XGBoost are used to predict the learning outcome. As per prior literature, Random Forest outperformed the other two (Rogers et al., 2025). When Random Forest is used with Synthetic Minority Oversampling Technique (SMOTE), it predicts the pass and fail students in the class, and this helps learners to work proactively (Jawad et al., 2022). Another study applies Artificial Neural Network, Logistic Regression, Naïve Bayes, and Decision Tree are used to predict the poor performance in computer science courses, where ANN outperforms the other approaches (Altabrawee et al., 2019).

Nowadays researchers utilize DL techniques on learning analytics to predict the students' learning outcome. DL techniques are capable of finding insights from raw data. Waheed et al predict the students' performance using deep artificial neural network, logistic regression and SVM, where deep artificial neural network outperforms with 84% - 93% accuracy (Waheed et al., 2020). Liu et al (2023) applied LSTM and One-dimensional Convolutional Neural Network (1D-CNN) along with traditional ML approaches to predict the learning outcome, where Long Short-Term Memor (LSTM) outperform with 90.25% accuracy. However, the application of DL to Learning analytics is still in the preliminary stages (Aljohani et al., 2019). Further, only a very few studies have applied the DL techniques to analyse and evaluate the learning behaviours in VLE (Aljohani et al., 2019).

3. Methodology

This study aims to predict the learner's performance using their interaction with VLE using ML and DL techniques.

3.1 Data

This study utilises two data sets, VLE log data and students' marks data. The VLE log data is collected from an asynchronous online course module offered in distance mode to external degree undergraduate students from a state university in Sri Lanka. The module was offered during the second year first semester of the degree programme. The module is followed by two hundred and forty-seven (247) students. The data contains 78,175 VLE click event records that are generated by the course administrators, instructors and learners. The course module discusses fourteen (14) topics using different resources. The module contains resources such as lecture materials and activities such as quizzes, forum discussions and assignments. The module was offered from August 2022 to December 2022. The module contains twenty (20) lessons, six (06) quizzes, and seven (07) activities. Table 1 shows the VLE log data attribute with description.

In addition to VLE log data, researchers utilise marks data which contains user, and the marks obtained by students.

Table 1: VLE Log data attributes and description

Attribute	Descriptions
Time	Timestamp of each interaction
User Full Name	User
Affected User	Other users involved in
Event Context	Resources or activity type
Component	Learning objectives
Event Name	Specific activity the user performs
Description	User event description
Origin	Client source
IP address	Users' IP address

3.2 Preprocessing

As a first step, the researchers removed the click events that are not generated by the students. The researchers removed 5,216 records generated by the instructors and administrators of the course module. The students generated 72,950 records used for the study. In the next step, two datasets were merged to generate a single file. Researchers removed the user's full name (for anonymisation) and replaced it with a unique ID to uniquely identify each student. Next, the time attribute is converted into a date and time format. Following that, the available numerical features were normalised using StandardScaler. In the next step, relevant attributes were extracted, and irrelevant features such as the IP address and origin were removed. At the subsequent step, the following details were extracted: (a) number of interactions per user, (b) weekday and weekend interactions, (c) extracted interactions per unique event type, (d) temporal features (average time and day of the interactions) and (e) extracted interactions per unique event type.

3.3 Method

After pre-processing, the study utilises the students' seven (7) behavioural or clickstream and thirty-four (34) learning activity features. Table 2 lists the behavioural or clickstream features used for the study with a description. Table 3 lists the learning activity features used by the study. The study utilises eighty (80) percent of the data for training and twenty (20) percent of the data for testing.

Table 2: Behavioural or Clickstream Features

Attribute	Description
unique_components	The number of distinct VLE components (e.g., quizzes, forums, videos) a student interacts with.
avg_hour	The average hour of the day when a student engages with the VLE.
std_hour	The standard deviation of the hours at which the student accesses the platform, indicating variability in study time.
avg_day	The average number of days per week that the student interacts with the VLE.
weekend_interactions	The total number of interactions occurring on weekends.
session_count	The total number of study sessions recorded for the student.
Peak study hour	The most frequent hour of the day when the student is active on the platform.

Table 3: Learning Activity Features

A file has been uploaded.	Quiz attempt updated
Comment created	Quiz attempt viewed
Comment deleted	The status of the submission has been viewed.
Course module instance list viewed	Tour started
Course module viewed	User list viewed
Course user report viewed	User profile viewed
Course viewed	User report viewed
Discussion created	Video module viewed
Discussion subscription deleted	Read tracking disabled
Discussion viewed	Remove submission confirmation viewed.
Feedback viewed	Role assigned
Grade overview report viewed	Some content has been posted.
Grade user report viewed	Submission created.
Post updated	Submission form viewed.
Quiz attempt auto saved	Submission updated.
Quiz attempt reviewed	Subscription created
Quiz attempt started	Quiz attempt summary viewed

The study utilises several prediction models, namely (a) traditional machine learning approaches: Random Forest Regressor (RF) and Support Vector Machines (SVM) and (b) Deep Learning Approaches: Multilayer Perceptron (MLP) and Recurrent Neural Networks (RNN). The RF model builds multiple decision trees and merges their prediction. The model used hundreds of trees to reduce the overfitting by using bootstrap aggregation. SVM uses high high-dimensional space to separate the classes. This study uses the Radial Basis Function as a kernel. The MLP model used hidden layers, batch normalisation and dropout to overcome overfitting. The model used the Adam optimiser. LSTM is a special RNN introduced to overcome the vanishing gradient problem. This study used Long Short-Term Memory (LSTM) with the Adam optimiser. The traditional ML approach RF is selected as the baseline for the study based on the literature (Borna et al., 2024). The study performed the result prediction with all the features after removing the absent student data. Further, the researchers performed the Exploratory Factor Analysis (EFA) to select the most influencing features and reduce the redundancy.

4. Results and Analysis

This section discusses the results obtained from the descriptive analysis and the model. Figure 1 shows the number of hits performed by different students. This shows that some students engaged more with the VLE while they were following the course. On the other hand, some did not. The maximum number of hits performed by students is 1,093, minimum number of hits perform by the student is two (2) and on average, students perform 295 hits. Figure 2 shows the number of hits performed on different activities. The Figure shows the system with the most influential activity, followed by a quiz. The very first place that the students enter is the system. The next most influencing activity is quiz, which shows that learners move more towards the path to growth.

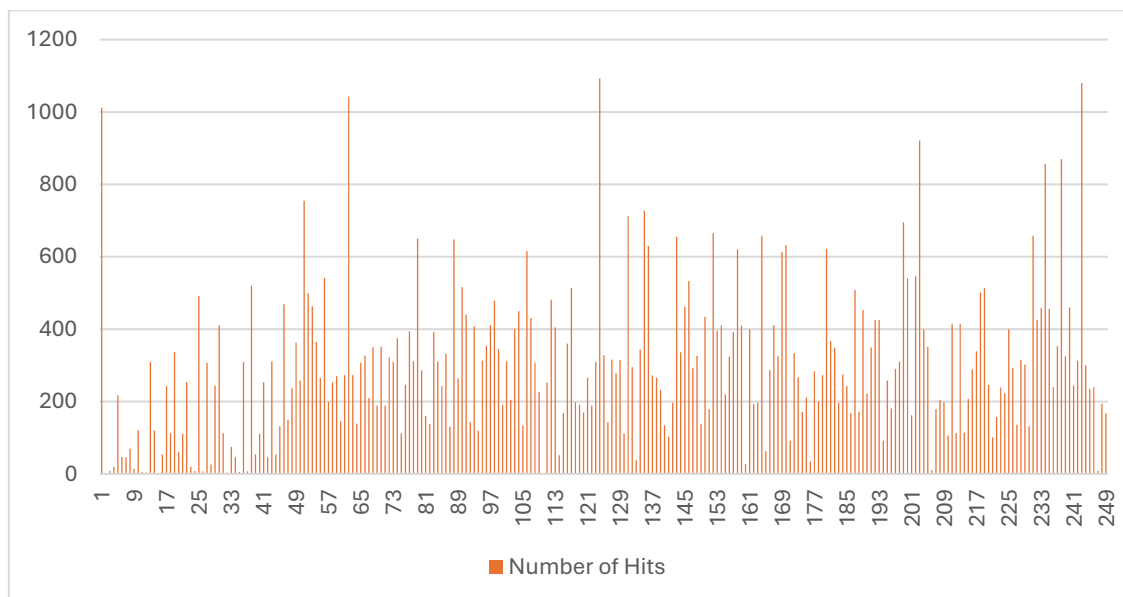


Figure 1: Number of hits by the students

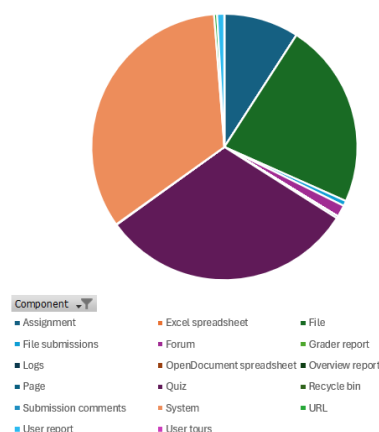


Figure 2: Different activities offered by the module and interaction

Figure 3 shows the number of events by hour of the day. This shows a higher number of interactions happening at night, ten (10) to eleven (11), and day ten (10) to eleven (11). Further, most of the user activity is happening in the evening from 6.00 pm to 12 midnight. This shows that students engaged more in their learning activities during late evening. This is shown in Figure 4. Figure 5 shows the activity distribution by the day of the week, where Wednesday is the most active day, followed by Saturday, Sunday Tuesday, Friday, Thursday and Monday. Weekdays have more interactions (70.3%) than on weekends (29.7%). Figure 6 shows the transition probabilities between the events. This Figure shows the highest transition probability between the quiz attempt submitted and the user graded.

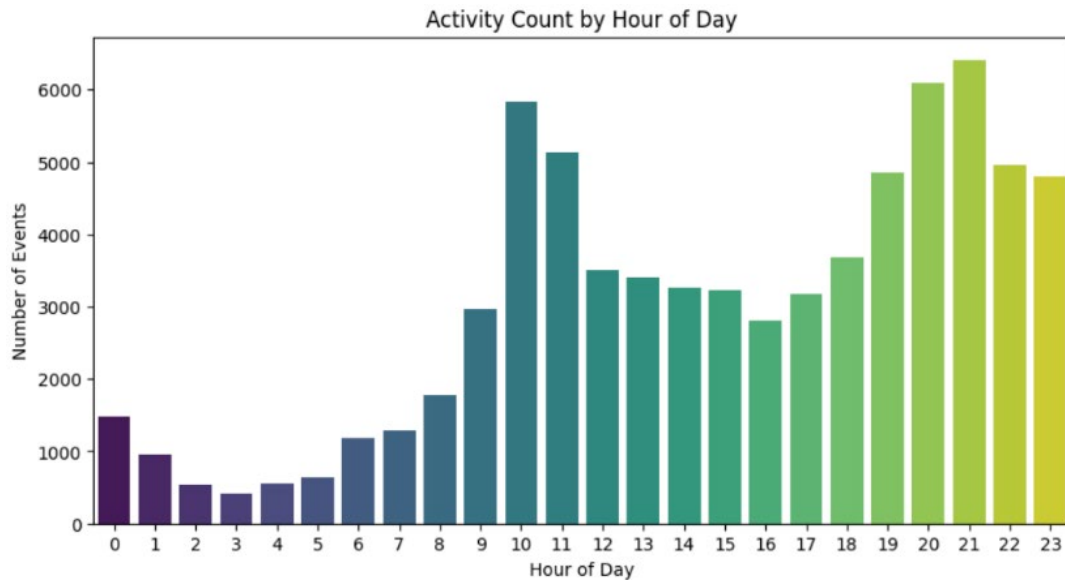


Figure 3: Activity count by hour of the day

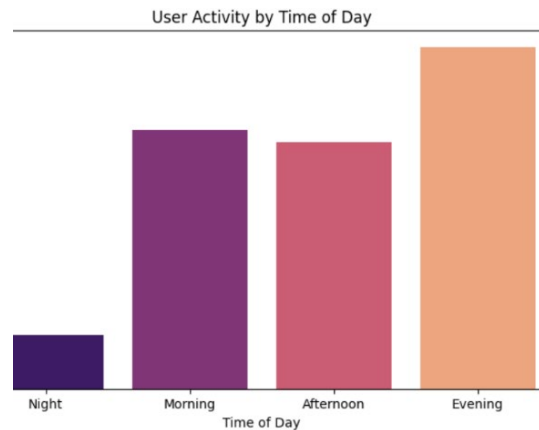


Figure 4: Activity by time of the day

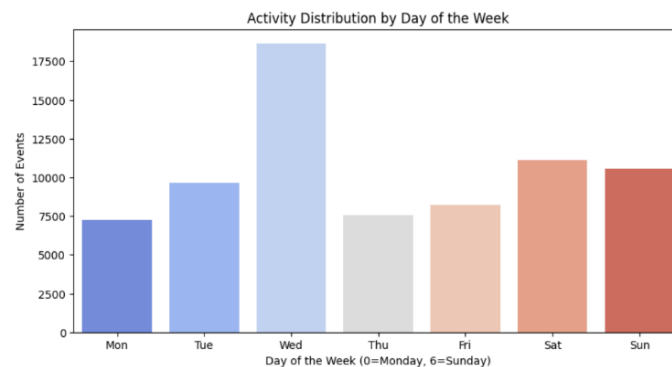


Figure 5: Activity distribution by day of the week

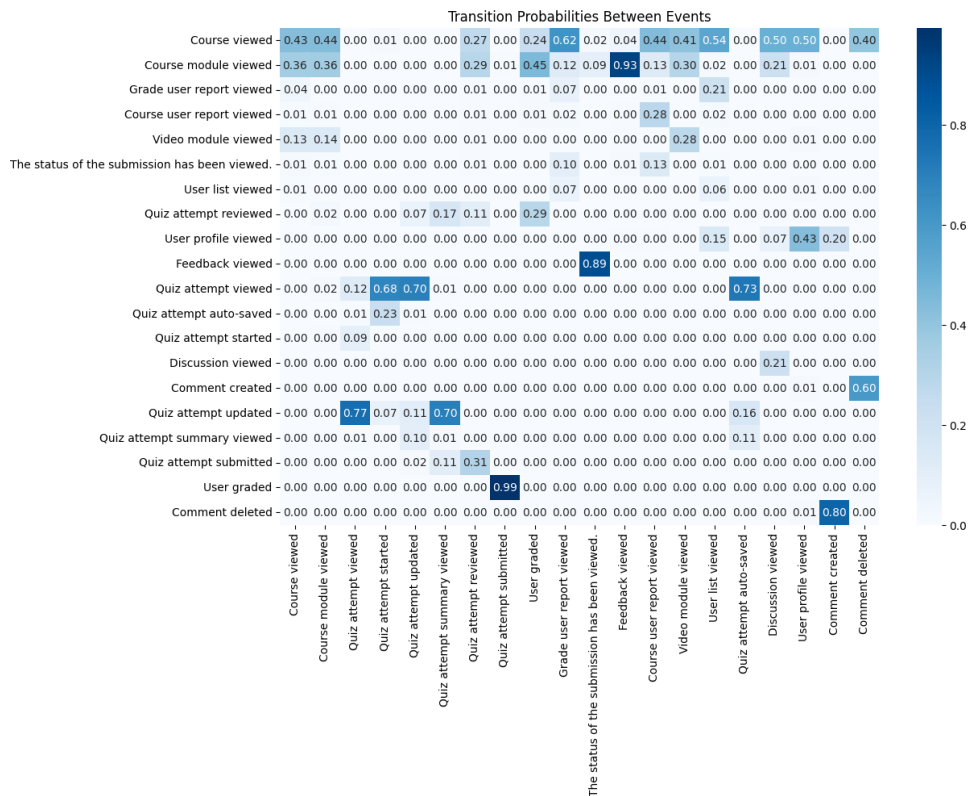


Figure 6: Transition probabilities between the events

The Model performance was evaluated using accuracy, F1-score and Mean Absolute Error (MAE). Table 4 shows the F1 score for different techniques where LSTM outperforms in terms of F1 score for pass and fail classes. Further, Table 5 lists the accuracy and Mean Absolute Error (MAE). The best model was the LSTM with 97% accuracy with 0.062 MAE. Even though MLP archived good accuracy, it failed to predict the failure class correctly.

Table 4: Approaches with F1 score

Technique	Class	Precision	Recall	F1 score
MLP	Pass	0.97	1.00	0.99
	Fail	0.00	0.00	0.00
RNN (LSTM)	Pass	1.00	0.97	0.99
	Fail	0.75	1.00	0.86
Random Forest	Pass	0.95	1.00	0.97
	Fail	1.00	0.33	0.50
SVM	Pass	1.00	0.58	0.74
	Fail	0.17	1.00	0.29

Table 5: Approaches with accuracy and Mean Absolute Error (MAE)

Technique	Accuracy	MAE
MLP	0.9744	0.0640
RNN (LSTM)	0.9744	0.0620
Random Forest	0.95	0.0512
SVM	0.62	0.051

Feature importance was analysed using the EFA. From the EFA, the most influential behavioural or clickstream features are the number of sessions count, number of weekend interactions and the number of unique

components. The most influencing learning activities are grade user report viewed, discussion created, discussion viewed, course viewed, a file has been uploaded, feedback viewed, course module viewed.

5. Discussion

The study analyses the learning behaviours of the learners along with their learning outcomes. It uses the number of hits to analyse the learner's engagement since the number of hits and hours spent on the activities have similar patterns (M.S. Faathima Fayaza & Supunmali Ahangama, 2024). The results show that students have more interaction on weekdays than on weekends. Furthermore, students engage more at night than during the day. The highest transition probability is between the "quiz attempt submitted" and the "user graded". The second highest transition probability is shown between "course module view" and "feedback viewed". After that, "the status of the submission has been viewed" and "feedback viewed" have the highest transition. These show that students expect feedback immediately when they perform their activity. Effective feedback motivates and encourages students to plan and monitor their learning strategies.

The study investigates several ML and DL approaches to predict the performance. The RF is used as the baseline model. However, the LSTM outperformed other approaches, possibly because the LSTM can handle sequential observations. Similar results have been reported by other researchers as well. For example, Aljohani et al. and Liu et al., reported that LSTM outperformed the ML approaches. (Aljohani et al., 2019; Liu et al., 2023)

Further, the study elaborates on the most influential behavioural and learning activity features in the performance prediction among distance learning students. The number of sessions, number of weekend interactions, number of unique components, grade user report viewed, discussion created, discussion viewed, course viewed, a file has been uploaded, feedback viewed and course module viewed are the most influential features in the performance prediction. This shows the active involvement of the students' impact on the learning outcomes, and these findings can be used to design and create the course module.

However, this study has some limitations. Clickstream data contains non-continuous data and results in the sparse data. The study used the data with the imbalance sample. This can be address in the feature by using the SMOTCK techniques. Future studies need to focus on predicting the performance in the granular grade level. Further, future research can also focus on the multimodal data such as assignment interaction and quiz to analyse the learning pattern.

6. Conclusion

Online learning provides flexibility in terms of time and place though the system suffers from a high dropout ratio and low retention rate. VLEs have become an increasingly popular platform for delivering online courses. VLE is designed as course centric. This study utilises the behavioural or clickstream features and learning activity features to predict and analyse the students' performance. The Random Forest Regressor, SVM (Traditional Machine Learning Approaches) and MLP and LSTM (Deep Learning Approaches) are used for performance prediction. Among these approaches LSTM outperforms with 97% accuracy. From the Exploratory Factor Analysis (EFA) session count, the number of weekend interactions, and the unique components are selected as the most influential behavioural features. Grade user report viewed, discussion created, discussion viewed, course viewed, a file has been uploaded, feedback viewed, and course module viewed have been selected as the most influential learning activity features.

Ethics Declaration

This study obtains ethical clearing from the University of Moratuwa for the data used in the study.

AI Declaration

This paper does not use any AI tools in its development.

References

- Alasalmi, T. (2021). Students Expectations on Learning Analytics: Learning Platform Features Supporting Self-regulated Learning. *International Conference on Computer Supported Education, CSEDU - Proceedings*, 2, 131–140. <https://doi.org/10.5220/0010537101310140>
- Aljohani, N. R., Fayoumi, A., & Hassan, S. U. (2019). Predicting at-risk students using clickstream data in the virtual learning environment. *Sustainability (Switzerland)*, 11(24). <https://doi.org/10.3390/su11247238>
- Altabrawee, H., Ali, O. A. J., & Ajmi, S. Q. (2019). Predicting Students' Performance Using Machine Learning Techniques. *JOURNAL OF UNIVERSITY OF BABYLON for Pure and Applied Sciences*, 27(1), 194–205. <https://doi.org/10.29196/jubpas.v27i1.2108>
- Al-Tameemi, G., Xue, J., Ajit, S., Kanakis, T., Hadi, I., Baker, T., Al-Khafajiy, M., & Al-Jumeily, R. (n.d.). *A Deep Neural Network-Based Prediction Model for Students' Academic Performance*.
- Borna, M. R., Saadat, H., Hojjati, A. T., & Akbari, E. (2024). Analyzing click data with AI: implications for student performance prediction and learning assessment. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1421479>
- Fazil, M., Rísquez, A., & Halpin, C. (2024). A Novel Deep Learning Model for Student Performance Prediction Using Engagement Data. *Journal of Learning Analytics*, 11(2), 23–41. <https://doi.org/10.18608/jla.2024.7985>
- Jawad, K., Shah, M. A., & Tahir, M. (2022). Students' Academic Performance and Engagement Prediction in a Virtual Learning Environment Using Random Forest with Data Balancing. *Sustainability (Switzerland)*, 14(22). <https://doi.org/10.3390/su142214795>
- Jo, Y., Maki, K., & Tomar, G. (2018). *Time Series Analysis of Clickstream Logs from Online Courses*. <http://arxiv.org/abs/1809.04177>
- Liu, Y., Fan, S., Xu, S., Sajjanhar, A., Yeom, S., & Wei, Y. (2023). Predicting Student Performance Using Clickstream Data and Machine Learning. *Education Sciences*, 13(1). <https://doi.org/10.3390/educsci13010017>
- Maaliw, R. R. (2021). A Personalized Virtual Learning Environment Using Multiple Modeling Techniques. *2021 IEEE 12th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2021*, 8–15. <https://doi.org/10.1109/UEMCON53757.2021.9666645>
- M.S. Faathima Fayaza & Supunmali Ahangama. (2024). *2024 4th International Conference on Advanced Research in Computing*.
- Pinner, R. S. (n.d.). *VLE or LMS: Taxonomy for Online Learning Environments*. www.iatefl.org
- Rizwan, S., Nee, C. K. & Garfan, S. (2025). Identifying the Factors Affecting Student Academic Performance and Engagement Prediction in MOOC using Deep Learning: A Systematic Literature Review. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3533915>
- Rogers, J. K., Mercado, T. C., & Cheng, R. (2025). Predicting student performance using Moodle data and machine learning with feature importance. *Indonesian Journal of Electrical Engineering and Computer Science*, 37(1), 223–231. <https://doi.org/10.11591/ijeecs.v37.i1.pp223-231>
- Ryani Kusumawati, R. (2024a). Leveraging Artificial Neural Networks to Predict and Enhance Student Performance in Virtual Learning Environments. *Researcher Academy Innovation Data Analysis*, 1(2), 148–159. <https://doi.org/10.69725/raida.v1i2.162>
- Ryani Kusumawati, R. (2024b). Leveraging Artificial Neural Networks to Predict and Enhance Student Performance in Virtual Learning Environments. *Researcher Academy Innovation Data Analysis*, 1(2), 148–159. <https://doi.org/10.69725/raida.v1i2.162>
- 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. <https://doi.org/10.1016/j.chb.2019.106189>
- Yağcı, M. (2022a). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1). <https://doi.org/10.1186/s40561-022-00192-z>
- Yağcı, M. (2022b). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1). <https://doi.org/10.1186/s40561-022-00192-z>
- Zakir, S., Hoque, M. E., Susanto, P., Nisaa, V., Alam, M. K., Khatimah, H., & Mulyani, E. (2025). Digital literacy and academic performance: the mediating roles of digital informal learning, self-efficacy, and students' digital competence. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1590274>