East African Journal of Information Technology
eajit.eanso.org
Volume 7, Issue 1, 2024
Print ISSN: 2707-5346 | Online ISSN: 2707-5354
Title DOI: https://doi.org/10.37284/2707-5354

EAST AFRICAN NATURE & SCIENCE ORGANIZATION

Original Article

Comparative Analysis of Machine Learning Classifier Models for Predicting Student Cognitive Load and Performance Outcomes in Moodle Learning Environment

Thaddeus Matundura Ogwoka[1], Prof. Robert Obwocha Oboko, PhD[2] & Prof. Christopher Kipchumba Chepken, PhD[2]*

[1] Technical University of Mombasa, P. O. Box 90420-80100, Mombasa, Kenya.
[2] University of Nairobi, P. O. Box 30197-00100. Nairobi, Kenya.
* Corresponding Author: ORCID ID: https://orcid.org/0009-0007-9935-6654; Email: togwoka@tum.ac.ke

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

Date Published:

20 September 2024

Keywords:

Learning Analytics, Moodle, Cognitive Load, E-Learning, Machine Learning.

ABSTRACT

Background: Advancements in ICT have driven the widespread adoption of e-learning platforms like Moodle, enhancing education delivery and experience. While research has primarily focused on predicting learner performance using Moodle data, the role of cognitive load has been underexplored. Purpose: This study aimed to compare machine learning classifiers in predicting student cognitive load and performance within Moodle. Methodology: Conducted at the Technical University of Mombasa in Kenya between November and December 2023, the experiment involved 415 undergraduate students in a four-week online course powered by Moodle LMS. Participants were randomly assigned to a treatment group containing cognitive load management interventions or a control group, without interventions. Using Moodle logs, the predictive performance of Naïve Bayes (NB), Random Forests (RF), Support Vector Machines (SVM), and K-Nearest Neighbours (KNN) was analyzed with Python's Jupyter package. Findings: Results showed that NB had high precision (96.96%) and accuracy (93.04%) with minimal training time, while SVM also performed well with higher training time. RF excelled in accuracy but required more computational resources. Conclusion: The study suggests NB and SVM effectively predict cognitive load and performance. This knowledge can be utilized by LMS and instructional designers to advance data-driven student interventions in supporting student success in e-learning environments.

APA CITATION

Ogwoka, T. M., Oboko, R. O. & Chepken, C. K. (2024). Comparative Analysis of Machine Learning Classifier Models for Predicting Student Cognitive Load and Performance Outcomes in Moodle Learning Environment. *East African Journal of Information Technology*, *7*(1), 301-317. https://doi.org/10.37284/eajit.7.1.2227

CHICAGO CITATION

Ogwoka, Thaddeus Matundura, Robert Obwocha Oboko and Christopher Kipchumba Chepken. 2024. "Comparative Analysis of Machine Learning Classifier Models for Predicting Student Cognitive Load and Performance Outcomes in Moodle Learning Environment". *East African Journal of Information Technology* 7 (1), 301-317. https://doi.org/10.37284/eajit.7.1.2227.

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

HARVARD CITATION

Ogwoka, T. M., Oboko, R. O. & Chepken, C. K. (2024) “Comparative Analysis of Machine Learning Classifier Models for Predicting Student Cognitive Load and Performance Outcomes in Moodle Learning Environment”, *East African Journal of Information Technology*, 7(1), pp. 301-317. doi: 10.37284/eajit.7.1.2227.

IEEE CITATION

T. M. Ogwoka, R. O. Oboko & C. K. Chepken “Comparative Analysis of Machine Learning Classifier Models for Predicting Student Cognitive Load and Performance Outcomes in Moodle Learning Environment.”, *EAJIT*, vol. 7, no. 1, pp. 301-317, Sep. 2024.

MLA CITATION

Ogwoka, Thaddeus Matundura, Robert Obwocha Oboko & Christopher Kipchumba Chepken “Comparative Analysis of Machine Learning Classifier Models for Predicting Student Cognitive Load and Performance Outcomes in Moodle Learning Environment.”. *East African Journal of Information Technology*, Vol. 7, no. 1, Sep. 2024, pp. 301-317, doi:10.37284/eajit.7.1.2227.

INTRODUCTION

This section sets the stage for the study and provides the background, objective, and significance of this research paper.

What is already known about this topic?

E-learning platforms like Moodle have transformed education by offering detailed data on student interactions, engagement, and performance (Krishnan et al., 2022). Past studies suggest that it is not the technology itself that determines student success on e-learning platforms, but rather how effectively it is used to positively influence learner cognitive processes (Kalyuga & Liu, 2015). Existing research on e-learning, particularly using Moodle, has primarily focused on predicting learner performance based on activity logs and learner profiles, without much focus on learner cognitive load that affects performance. Machine learning models, like Naïve Bayes, Random Forests, Support Vector Machines, and K-Nearest Neighbours, have shown effectiveness in predicting student success, yet comprehensive studies on cognitive load prediction are limited (Romero & Ventura, 2020).

What does this paper add?

The experimental results from this study, of predicting student cognitive load and performance within Moodle highlight the effectiveness of integrating algorithm models into data-driven interventions for enhancing e-learning outcomes. This paper provides a comparative analysis of these algorithm models, demonstrating their value in improving instructional design and learner support within e-learning environments.

Background

In recent years, there has been widespread adoption of e-learning in education. This gradual increase in the use of e-learning platforms, mostly powered by learning management systems (LMSs) such as Moodle, has enhanced the way education is delivered and experienced (Garkule & Makarevičs, 2018).

Moodle is a widely adopted open-source learning management system that captures valuable data log information on student interactions, engagement, and performance (Yen A. C et al., 2015). This platform offers a unique opportunity to understand and optimize the learning process of learners (Krishnan et al., 2022). However, the key to successful e-learning lies not solely in the technology itself, but rather in how it is utilized to positively impact learner activities and cognitive processes (Kalyuga & Liu, 2015).

Cognitive load is the amount of working memory resources required to process information during learning (Skulmowski, & Rey, 2017). This load can be categorized into intrinsic cognitive load, due to task complexity, extraneous cognitive load from irrelevant content presentation, and germane cognitive load from useful resources for schema creation (Sun et al., 2023; van Mierlo et al., 2012a).

In the Moodle learning environment, managing cognitive load is essential for optimizing educational outcomes (Garkule & Makarevičs, 2018). According to John Anderson’s theory of information processing, learners’ information processing capacity is limited, and exceeding this

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

capacity can disorient learners and negatively impact their learning experiences (Kalyuga & Liu, 2015).

İn e-learning environment, cognitive load can be measured by how deeply students interact with the learning content, as indicated by Moodle's learner behavioural data logs (A. C. Yen et al., 2015). These data logs can be mapped to cognitive load and performance indicators, providing insights into learner cognitive depth and engagement (D. R. Garrison et al., 1999; *Learning Analytics Indicators - MoodleDocs*, n.d.). For example, if some learners view activities more frequently than others, it may suggest a higher cognitive load due to either extraneous cognitive load from irrelevant material or intrinsic cognitive load from complex tasks (van Mierlo et al., 2012b; C.-H. Yen et al., 2015).

Predictive modelling of cognitive load and learner performance using Moodle data holds great potential for enhancing educational outcomes and facilitating personalized learning (Romero & Ventura, 2020). Machine learning techniques can identify at-risk students early, enabling timely interventions to improve their outcomes (Lauren & Vanessa, 2022; Oppong, 2023).

However, despite the significant potential of machine learning in educational settings such as in Moodle learning environment (Hasan et al., 2020), limited comprehensive studies exist, specifically comparing different machine learning classifier models for effective prediction of student cognitive load and performance outcome.

Objective of the Study

This study aimed to conduct a comparative analysis of machine learning classifier models for predicting student cognitive load and performance outcomes within the Moodle learning environment.

To achieve this objective, the following research question was formulated for this investigation:

> *Which machine learning classifier models are most effective for predicting student cognitive load and performance outcomes within the Moodle learning environment?*

Significance of the Study

The study's findings underscore the importance of integrating machine learning algorithms into instructional and LMS designs to enhance e-learning effectiveness. This integration provides critical insights into learners' cognitive load and performance, enabling timely support interventions that can significantly improve learning outcomes. Practically, it informs the design of data-driven strategies aimed at boosting student success in e-learning environments. Policymakers can leverage these insights to develop robust frameworks that incorporate cognitive load management into online education, ensuring a more tailored and effective learning experience.

LITERATURE REVIEW

This section offers a survey of literature to get a preliminary understanding of the existing research related to predicting student cognitive load and performance utilizing machine learning in e-learning environments, particularly using the Moodle learning management system.

Theoretical Review

Learning is the process of acquiring new or modifying existing knowledge, resulting in behaviour change and performance improvement (Young, 2015). Whether through physical interaction or an e-learning interface, the fundamental goal of learning is knowledge acquisition (Schunk, 2012).

E-learning platforms, powered by LMSs like Moodle, have revolutionized education delivery (Garkule & Makarevičs, 2018). These platforms collect extensive data on student interactions, engagement, and behaviour, which can be analysed to gain insights into learning outcomes and provide timely interventions for struggling students (Segura et al., 2022). Additionally, it is important to consider that students’ working memory is limited (Sucharitha et al., 2020).

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

Various philosophies explain how learning occurs, informing the design of e-learning systems to align with human learning behaviour. Cognitivism, for example, focuses on changes in mental structures and memory storage during learning (Hoque, 2016, Song, & Thompson, 2011). Cognitive load theory, introduced by John Sweller, posits that the mental effort required to process information significantly affects learning. The cognitive load consists of intrinsic load (due to the complexity of the content), extraneous load (due to the way content is presented), and germane load (for schema creation) (Kalyuga, & Liu, 2015). Optimal cognitive load is crucial for effective learning, as excessive load can hinder learning, while appropriate levels enhance comprehension and retention (Kante et al., 2016).

E-learning interfaces should consider learners' cognitive processes to avoid overloading their working memory, especially in e-learning where learners engage in self-paced learning (Aeiad & Meziane, 2019, pp. 1485–1486; Kolekar et al., 2018).

Analyzing Cognitive Load and Performance Indicators in Moodle Data

Analysing both cognitive load and performance outcomes in a Learning Management System like Moodle is essential, as learner performance is directly influenced by cognitive load (Sweller et al., 2011). Measuring cognitive load reveals critical information about instructional conditions and their impact on learner performance (Juhanie & Paas., 2017, p. 134).

Moodle LMS provides various metrics for tracking and evaluating learning effectiveness based on learner behaviours (Młynarska et al., 2016). These metrics include learner logins and activity levels, forum participation, course completion rates, assessment and quiz performance, and navigation patterns (Dobre, 2015, Romero & Ventura, 2020). Collectively, these metrics provide a comprehensive view of learner engagement, progress, and course effectiveness (A. C. Yen et al., 2015).

Empirical literature indicates that learner activities captured in Moodle logs serve as indicators for cognitive load constructs (Sun et al., 2023). The variation in the count of learner interactions with such indicators can be mapped to the cognitive load level of learners, whose impact can be measured by their performance (A. C. Yen et al., 2015). By analyzing these logs, researchers can correlate learner behaviour with cognitive load measures using the element interactivity of the learning material (Kalyuga, 2012). Therefore, data analytics on learner behaviour through log activities in Moodle, reveal patterns that can indicate learner cognitive load (C.-H. Yen et al., 2015).

Predicting Cognitive Load and Performance in Moodle through Machine Learning

Research has explored the predictive modelling of student cognitive load and performance in Moodle learning environments (Sun et al., 2023). Modeling in this context involves creating predictive models that use machine learning algorithms to identify and address the factors contributing to cognitive load (Krishnan et al., 2022; Tlili et al., 2023).

Examples of machine learning algorithms commonly used in e-learning for predictive modelling include decision trees, support vector machines, and neural networks (Boulesteix, & Strobl, 2009). This is an indication of the significant potential of machine learning applications in educational settings. Research is still ongoing in these endeavours (Romero & Ventura, 2020).

For example, Shayan, & van Zaanen (2019), employed Decision Trees and JD3 algorithms on Moodle LMS data logs to distinguish between weak and strong students. Utilizing Random Forest, Gradient Boosting, K-nearest Neighbours, and Linear Discriminant Analysis algorithms, Quinn, & Gray (2019), investigated whether data from Moodle LMS can be used to predict student performance outcomes by constructing measures of learner activities from Moodle data logs. Quinn

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

& Gray's results indicated that classifiers predicted student grades moderately well.

In "Predicting Student Satisfaction of Emergency Remote Learning in Higher Education during COVID-19", Ho et al., (2021) employed KNN, SVR, MLPR, LightGBM, RF, and ENet regression machine learning models to examine significant predictors influencing undergraduate student satisfaction.

Kaensar, & Wongnin (2023), conducted an analysis and prediction of student performance based on Moodle log data using a Support Vector Machine, Neural Network, Random Forest, Decision Tree, Logistic Regression, and Linear Regression machine learning techniques to forecast student performance.

Overall, these studies collectively underscore the importance of predictive analytics modelling to understand and improve student outcomes in LMS environments. However, the majority of these studies have modelled learner performance, yet cognitive load influences performance.

Table 1 provides a comparison of some of these studies, illustrating the Learning Management System (LMS) and tools utilized, as well as whether they predicted both performance and cognitive load. From these studies, Moodle is widely employed, while the majority of studies predict performance, neglecting the influence of cognitive load, which significantly impacts performance.

Table 1. Comparing Various Studies on Modeling Cognitive Load and Performance

Author(s)	LMS Used	Tool(s) Used	Performance	Cognitive Load
Shayan & van Zaanen(2019)	Moodle	WEKA, R	✓	x
Quinn & Gray (2019)	Moodle	R	✓	x
Hasan et al. (2020)	Moodle	Orange	✓	x
Suad et al. (2023)	Moodle	Moodle LA	✓	x
Ho et al. (2021)	Moodle	Python	✓	X
Ayouni et al. (2021)	Blackboard	Not mentioned	✓	x
Sun et al. (2023)	Notmentioned	R	x	✓
Conijn et al. (2017)	Moodle	R	✓	x

Research Gap on Predictive Analytics in Learning Management Systems

Despite the growing use of e-learning platforms like Moodle, there is limited research on utilizing traces of learner behaviours in e-learning to identify and explore variations in cognitive load and student engagement (Sun et al., 2023). There exist limited comprehensive studies for comparing different machine learning classifier models for their effectiveness in predicting student cognitive load and performance outcomes. Theng, & Theng (2020), aver, that among the "Predictive analytics problems that need in-depth research" in education settings.

METHODS

This section explains how the research design was conducted. Encompassing data collection, preprocessing, feature selection, classifier selection, model training, and performance evaluation.

Research Design

To answer this study's objective, an experiment was conducted at the Technical University of Mombasa in Kenya between November and December 2023, involving 415 undergraduate students who were randomly chosen to study an online short course using the university's Moodle LMS for four weeks. The study utilized purposeful sampling participants who had previously studied a web programming course

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

were requested to study the LA Ravel framework short course free of charge online. This method was chosen to ensure sufficient data logs for the study.

The participants were randomly enrolled in either a course with interventions aimed at managing their cognitive load (treatment group) or a similar course without interventions (control group). The course content was designed the same way, and then interventions were added to the treatment group. Data were collected through Moodle LMS activity logs, which were automatically recorded in the Moodle databases as the participants interacted with the LMS. The study adhered to the "Intention-to-treat" principle to maintain the validity of the results, analyzing participants in their assigned groups regardless of their study completion.

Data collection

After participants were randomly enrolled in a Moodle-based short course, Moodle automatically recorded detailed log data of learners. At the end of each week, these logs were collected by downloading the log file as a Comma comma-separated version (CSV) file from the Moodle LMS database.

The downloaded CSV Moodle data log file contained 9 fields: “Time”; describing the timestamp of the log record, “user full name”; for the student’s full name, ”affected user”; for the affected student’s full name, “event context”; representing the context of the student event, “component”; for component producing the log record, “event name”; for name of the student event, “description”; for describing the student event, “origin”; for describing the origin of the log record (client/webserver), and “IP address”; for describing the IP address of the device through which the student logged in (Rachel et al., 2018). Table 2 shows these fields of the downloaded Moodle raw log data.

Table 2. Moodle Data Log File showing the 9 Fields

Time	User full name	Affected user	Event context	Component	Event name	Description	Origin	IP address
2/06/23, 14:22	Dr. Kennedy Ha	-	Course: Advanced Web Programming Group 1	Logs	Log report viewed	The user with id '515' view	web	41.89.128.5
2/06/23, 14:21	Dr. Kennedy Ha	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '515' view	web	41.89.128.5
2/06/23, 14:21	Dr. Kennedy Ha	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '515' view	web	41.89.128.5
2/06/23, 14:21	Dr. Kennedy Ha	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '515' view	web	41.89.128.5
2/06/23, 14:20	Dr. Kennedy Ha	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '515' view	web	41.89.128.5
2/06/23, 13:47	Dr. Kennedy Ha	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '515' view	web	41.89.128.5
2/06/23, 13:46	Dr. Kennedy Ha	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '515' view	web	41.89.128.5
2/06/23, 13:45	Dr. Kennedy Ha	-	File: Lesson 1	File	Course module view	The user with id '515' view	web	41.89.128.5
2/06/23, 13:45	Dr. Kennedy Ha	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '515' view	web	41.89.128.5
2/06/23, 13:45	Dr. Kennedy Ha	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '515' view	web	41.89.128.5
2/06/23, 01:20	OENGA PETE	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '2928' vie	web	197.232.55.242
1/06/23, 09:40	NGEI LARRY	-	URL: Session 3 Web Conference	URL	Course module view	The user with id '3604' vie	web	197.232.26.117
1/06/23, 09:40	NGEI LARRY	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '3604' vie	web	197.232.26.117
1/06/23, 09:40	NGEI LARRY	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '3604' vie	web	197.232.26.117
1/06/23, 08:53	NGEI LARRY	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '3604' vie	web	197.232.26.117
1/06/23, 08:35	REAGAN ATH	-	URL: LESSON 2 RECORDED WEB CONFI	URL	Course module view	The user with id '938' view	web	41.90.217.241
1/06/23, 08:35	REAGAN ATH	-	File: Session 2: Routes, Controllers and Views	File	Course module view	The user with id '938' view	web	41.90.217.241
1/06/23, 08:34	REAGAN ATH	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '938' view	web	41.90.217.241
1/06/23, 08:34	AMIANI PAUL	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '940' view	web	197.181.114.170
1/06/23, 08:34	AMIANI PAUL	-	File: Session 4-Class Notes	File	Course module view	The user with id '940' view	web	197.181.114.170
1/06/23, 08:34	AMIANI PAUL	-	Course: Advanced Web Programming Group 1	System	Course viewed	The user with id '940' view	web	197.181.114.170

Data Preprocessing

The primary objective of pre-processing was to extract pertinent and accurate information for subsequent analysis (Rachel et al., 2018). This aimed to ensure that the collected data was clean, properly scaled, and split into cognitive load and performance indicator features, as well as cognitive load level and performance outcome labels.

Dataset Cleaning: The first step in cleaning the downloaded Moodle datasets was to extract the fields of interest from the nine fields in the downloaded file. Three fields of interest were identified and formatted using pivot tables in the Python Jupyter package (Rachel et al., 2018). Rows represented "User full name," columns represented "Event name," and values represented the "count of time." Next, duplicates and records of course administrators were removed, the

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

timestamp field was parsed, and missing values were addressed by arbitrary zero imputation since missing values indicated the absence of an event or activity by the learner. Finally, learner identities were anonymized. Table 3 shows the cleaned records.

Table 3. Cleaned Moodle Data Logs

student	Course module viewed	Course viewed	Quiz attempt started	Quiz attempt viewed	Quiz attempt submitted	Course activity completion updated	Quiz
std1	49	51	8	52	8	49	18
std2	6	2	0	0	0	7	0
std3	6	13	0	0	0	7	0
std4	52	61	4	18	4	31	0
std5	0	1	0	0	0	0	0
std6	58	48	3	20	3	33	0
std7	1	4	0	0	0	2	0
std8	16	8	0	10	0	19	0
std9	34	50	2	11	1	19	0
std10	7	9	0	0	0	7	0
std11	2	2	0	0	0	4	0
std12	6	19	0	6	0	10	0
std13	2	6	0	0	0	4	0
std14	56	78	5	46	4	47	15
std15	11	16	0	0	0	10	0
std16	2	7	0	0	0	4	0
std17	11	10	2	16	2	13	0
std18	45	43	0	23	0	44	0
std19	1	4	0	0	0	1	0
std20	32	53	0	10	0	26	0
std21	42	64	1	10	1	23	0

Scaling Numeric Learner Activities to Cognitive Depth and Performance Scale:

Moodle learning analytics provides a model for measuring learner cognitive presence through cognitive depth indicators based on learner activity type, ranging from 0 to 5 (D. Garrison et al., 2000; Learning Analytics Indicators - MoodleDocs, n.d.). According to Moodle's cognitive depth level model, all indicators for cognitive load were scaled from 0, indicating 0% cognitive participation, to 1, indicating 100% cognitive participation.

The scaling strategy used was: if a value was greater than or equal to half the maximum numeric value in the Moodle log for learner activity views, assign 1; otherwise, assign 0.

For performance outcome evaluation, Moodle's learning analytics model assigns:

- 0 to quiz performance indicators if a student did not view the activity or attempt the quiz,

- 1 if a student completed all the course activities or submitted the quiz,

- 2 if the student submitted and viewed feedback,

- 3 If feedback was provided to peers or instructors,

- 4 if the quiz/activity was revised and resubmitted (D. Garrison et al., 2000; Learning Analytics Indicators - MoodleDocs, n.d.)

In this study, two Moodle indicators "Activity completion updated" and "quiz" which constituted three assignments given to students totalling 30 marks, were scaled to 0 or 1 using the following criteria: For "activity completion updated," if the value was greater than or equal to 50, assign binary 1; otherwise, assign binary 0. For the "quiz" indicator, if the value was greater than or equal to 15, assign 1; otherwise, assign 0. The overall performance outcome was evaluated based on the condition: if either "activity completion updated" or "quiz" was 1, the "performance outcome" was labelled "Pass"; otherwise, it was labelled "Fail".

From the extracted indicators, correlation analysis was performed to assess the relationships between indicators for feature selection, and a correlation matrix was constructed. Strong positive correlations were observed, such as between "Course module viewed" and "Course viewed" (0.891), highlighting significant associations that validate the study's measurements. Figure 1 illustrates the correlation matrix plot for the treatment group, and Table 4 and Table 5 show the selected cognitive load and performance outcome

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

features respectively. Table 6 illustrates the mapping of the selected features to cognitive load and performance constructs.

Figure 1. **Treatment Group Correlation Matrix**

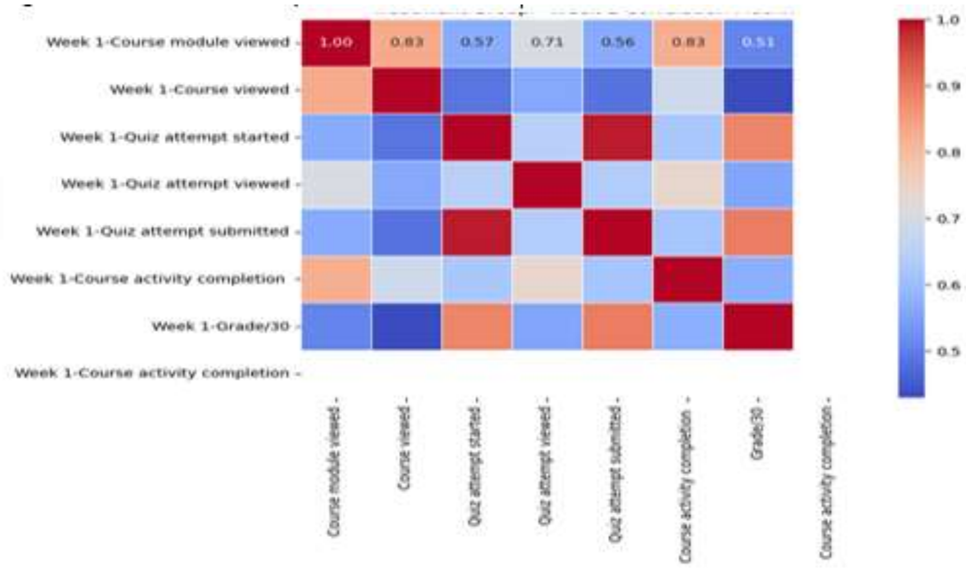


Table 4. Selected Cognitive Load Features

student	course module viewed:extraneous cognitive load	quiz attempt viewed:intrisic cognitive load	Quiz attempt submitted:germane cognitive load
std1	0	0	0
std2	1	1	1
std3	0	0	0
std4	1	1	1
std5	0	0	1
std6	0	0	0
std7	0	0	0
std8	0	0	1
std9	1	1	1
std10	0	0	0
std11	0	0	1
std12	0	0	0
std13	0	0	0
std14	0	0	0
std15	0	0	0
std16	0	0	0
std17	0	1	1
std18	0	0	0
std19	0	0	0
std20	0	0	0
std21	0	0	0
std22	1	0	1

Table 5. **Selected Performance Outcome Indicators**

student	Course activity completion updated	quiz/30	performance outcome
std1	0	1	Pass
std2	0	0	Fail
std3	0	0	Fail
std4	0	0	Fail
std5	0	0	Fail
std6	0	0	Fail
std7	0	0	Fail
std8	0	0	Fail
std9	0	0	Fail
std10	0	0	Fail
std11	0	0	Fail
std12	0	0	Fail
std13	0	0	Fail
std14	0	1	Pass
std15	0	0	Fail
std16	0	0	Fail
std17	0	0	Fail
std18	0	0	Fail
std19	0	0	Fail
std20	0	0	Fail
std21	0	0	Fail
std22	0	0	Fail

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

Table 6. Mapping Moodle Indicators to Cognitive Depth Levels and Performance Constructs

Moodle LMS indicator	Cognitive Depth Level	Cognitive Load Construct
Course module viewed	1: 0% (not viewed)- 0, 100%(viewed)- 1	Extraneous Cognitive Load
Course viewed		
Quiz attempt viewed	4: 20% (viewed/started)- 0.8, 40% (submitted)- 0.16	Intrinsic Cognitive Load
Quiz attempt Started		
Quiz attempt submitted		Germane Cognitive Load
Course activity completion	completed-1, not completed-0	Performance Construct
Quiz grades	pass:1, Fail:0	

The criteria for selecting the algorithms were based on the predictive problem and numeric nature of Moodle datasets collected (Geron, 2016). Based on these criteria, from the numerous existing algorithms used in data analytics problems, Naïve Bayes (NB), Random Forests (RF), Support Vector Machines (SVM), and K-Nearest Neighbours (KNN) were selected for this study. According to Theng & Theng's (2020) experimental literature, NB, SVM, KNN, and RF are widely used in predictive data analytics problems. Table 7 illustrates the features of some of these algorithms.

Table 7. Data Analytics Algorithm Features

Algorithm	Feature
Naïve Bayes (NB)	Highly scalable, and supports both continuous and discrete data.
SVM	Handles classification and regression of high-dimensional data, binary.
KNN	Effective for both classification and regression when dealing with small to moderately-sized datasets
Linear Regression	Great for predicting continuous numerical values
Decision Trees	Effective for both classification and regression tasks.
Random Forest	Multiple decision trees to improve prediction accuracy and reduce overfitting handle a wide range of data types.
Neural Networks	CNNs for images and RNNs for sequences-large amounts of data.

Model Training

In the training phase, the learning model was trained using transformed data (Theng & Theng, 2020). The training dataset was divided into three parts: training set to teach the model to make predictions, validation set, to help in tuning model parameters and selecting the best-performing model, and test set, to evaluate the model's final performance. Based on the size of the Moodle dataset pre-processed, the dataset was split into 80% for training, 10% for validation, and 10% for testing (80/10/10), the commonly used practice in research (Hstie et al., 2013).

Model Performance Evaluation

After training each model on the training set, performance was evaluated on the testing set based on training time, accuracy, precision, error, and area under curve metrics. Similar studies have used this approach (Theng & Theng, 2020; Wasylewicz & Scheepers-Hoeks, 2018).

RESULTS

The Results section is divided into two parts. The first part presents a preliminary descriptive analysis of the collected data. The second part discusses the comparative performance of

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

machine learning algorithms used in the study for predictive analytics.

Preliminary Descriptive Analysis Validity and Reliability

Reliability:

The author performed reliability testing using the factor analysis method. The component matrix provides information about the relationship between the original variables and the principal components, which can help in understanding the underlying structure of the data and identifying patterns or constructs represented by the components.

For the treatment group, all the indicators showed strong high loadings on Component 1 loading. İndicating high reliability among the variables. Figure 2 illustrates these results.

Figure 2. Control Group Data Reliability Results

Component Matrix[a]

	Component
	1
Course module viewed	.884
Course viewed	.792
Quiz attempt started	.845
Quiz attempt viewed	.844
Quiz attempt submitted	.841
Course activity completion	.884

week1

Component Matrix[a]

	Component
	1
Course module viewed	.905
Course viewed	.841
Quiz attempt started	.850
Quiz attempt viewed	.888
Quiz attempt submitted	.854
Course activity completion	.930

week2

Component Matrix[a]

	Component
	1
Course module viewed	.918
Course viewed	.863
Quiz attempt started	.830
Quiz attempt viewed	.889
Quiz attempt submitted	.841
Course activity completion	.919

week3

Component Matrix[a]

	Component
	1
Course module viewed	.908
Course viewed	.880
Quiz attempt started	.817
Quiz attempt viewed	.888
Quiz attempt submitted	.814
Course activity completion	.933

week4

Similarly, control group data showed strong reliability for the four weeks. Figure 3 illustrates these results.

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

Figure 3. Control Group Data Reliability Results

Component Matrix[a]

	Component
	1
Course module viewed	.866
Course viewed	.723
Quiz attempt started	.938
Quiz attempt viewed	.948
Quiz attempt submitted	.925
Course activity completion	.855

week1

Component Matrix[a]

	Component
	1
Course module viewed	.960
Course viewed	.895
Quiz attempt started	.977
Quiz attempt viewed	.966
Quiz attempt submitted	.966
Course activity completion	.953

week 2

Component Matrix[a]

	Component
	1
Course module viewed	.940
Course viewed	.861
Quiz attempt started	.965
Quiz attempt viewed	.955
Quiz attempt submitted	.955
Course activity completion	.928

week3

Component Matrix[a]

	Component
	1
Course viewed	.890
Quiz attempt started	.984
Quiz attempt viewed	.973
Quiz attempt submitted	.972
Course activity completion	.947

week4

Validity:

To measure the consistency of the instrument, Cronbach's alpha measure was used at a confidence level of 95%. Higher values of Cronbach's alpha (typically above 0.70) suggest greater internal consistency reliability.

The Cronbach's alpha values for both the treatment and control groups measured showed good internal item consistency, as illustrated in Table 8.

Table 8. Validity Test Results for Treatment and Control Groups

Week	**Cronbach's Alpha**	
	Treatment Group	Control Group
Week 1	(0.8056287867410907, [0.756, **0.848**])	(0.7442468012184887, [0.683, **0.797**])
Week 2	(0.8278156408131485, [0.786, **0.864**])	(0.859467445917776, [0.83, **0.885**])
Week 3	(0.8476885487280352, [0.812, **0.879**])	(0.8470930435532028, [0.814, **0.876**])
Week 4	(0.8443597784902894, [0.808, **0.876**])	(0.8592615203074347, [0.829, **0.886**])

Weekly Learner Activity Visualization Trends

On average, the treatment group learners' performance was moderately high compared to the control group. This is indicated in the course activity completion and Grade/30 performance indicators. The variations in the trends of the cognitive load indicators for the two groups suggest the effect of the interventions provided for the treatment group. Figure 4 and Figure 5, show these trends.

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

Figure 4. **Treatment Group Learner Weekly Activity Behaviours**

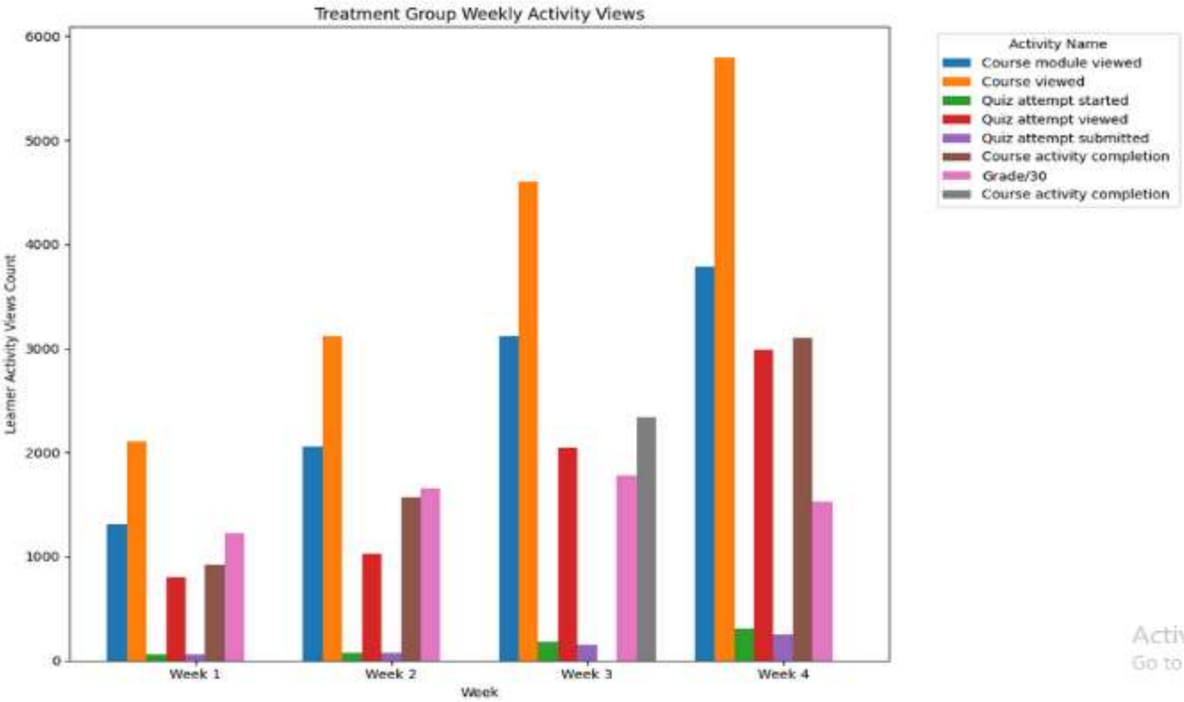
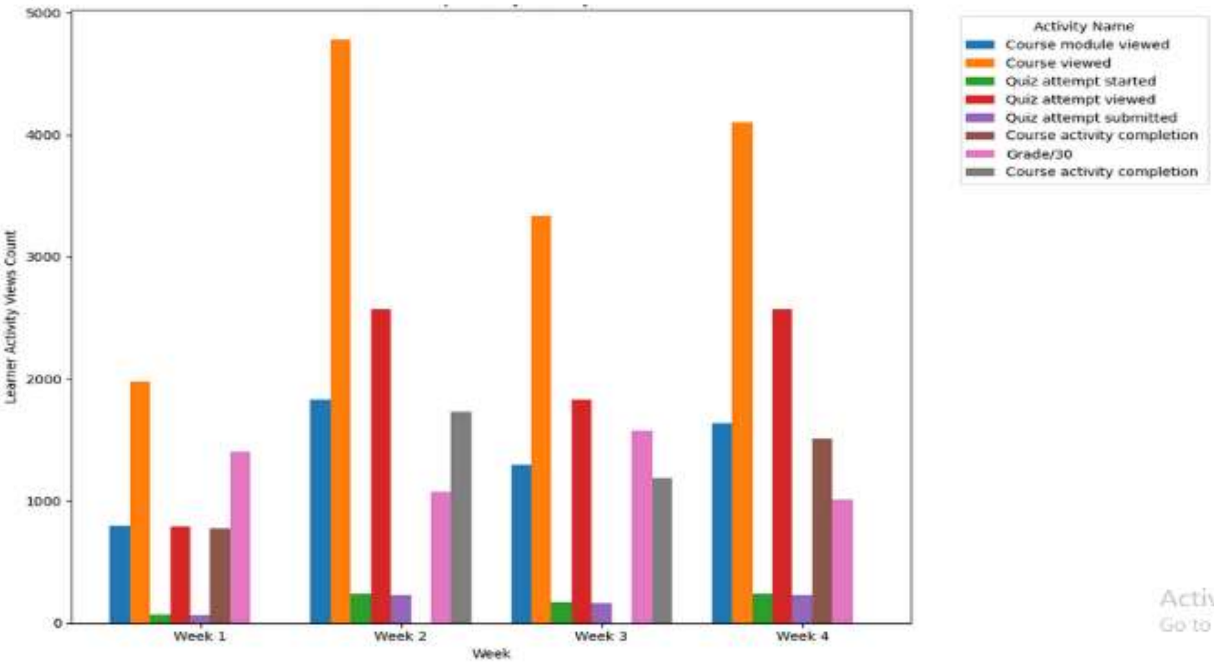


Figure 5. Control Group Learner Weekly Activity Behaviours



Comparative Performance of Machine Learning Algorithms

On precision and accuracy scores, the Naive Bayes (NB) model consistently performed well in predicting cognitive load and performance, achieving a high precision of 96.96% in the treatment group compared to 88.82% in the control group, and an accuracy of 93.04% in the treatment group compared to 92.32% in the control group with a minimal training time of 0.00 ms. Support Vector Machine (SVM) also demonstrated strong predictive capabilities, with precision values of 95.56% in the treatment group

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

and 87.63% in the control group, with accuracy values of 95.55% in the treatment group and 84.61% in the control group, though requiring a slightly longer training time of 15.61 ms in the treatment group, 0.00 ms in the control group.

K-Nearest Neighbours (KNN) showed competitive accuracy metrics, with precision values of 94.58% in the treatment group and 86.77% in the control group, and required moderate training time of 7.00 ms in the treatment group, and 0.00 ms in the control group. Random Forest had an intensive training time of 174.00 ms in the treatment group, and 153.63 ms in the control group but delivered reliable predictions with high accuracy and AUC.

The variations in the performance metrics between models and groups highlight their differing abilities to effectively capture and predict student outcomes. These differences can be attributed to the interventions received by the treatment group, which likely enhanced learner engagement and performance, leading to more consistent data variations in the treatment group learners, compared to the control group learners.

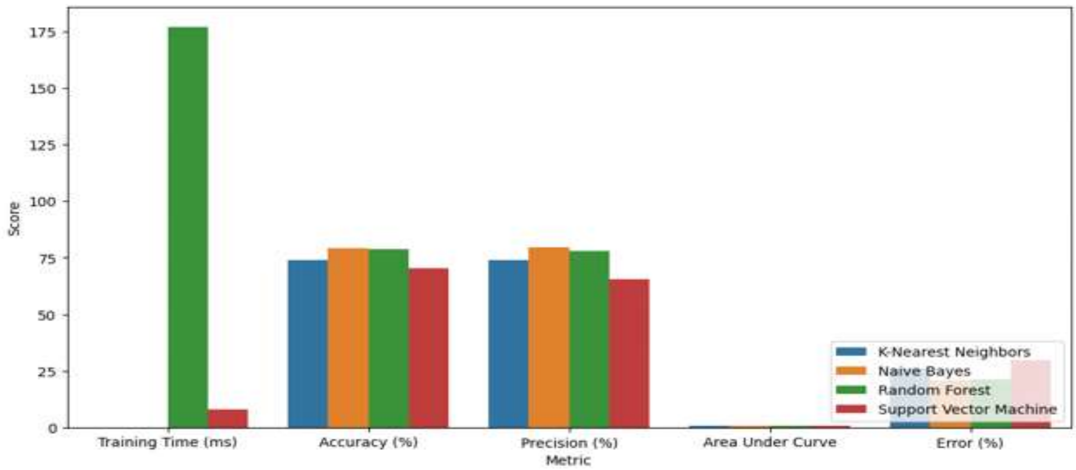
Which machine learning classifier models are most effective for predicting student cognitive load and performance outcomes within the Moodle learning environment?

In summary, Naive Bayes and SVM proved effective across both cognitive load and performance predictions, balancing predictive power with efficiency, while Random Forest excelled in accuracy at the cost of increased computational demands. These findings underscore the effectiveness of NB and SVM in balancing predictive power with efficiency, while RF excelled in accuracy despite higher computational demands. Table 9, Figure 6, Figure 7, Figure 8, and Figure 9 illustrate these results.

Table 9. Comparative Analysis of Prediction Algorithms

Treatment Group Datasets								
Performance Metrics	Cognitive Load				Performance			
	NB	SVM	KNN	RF	NB	SVM	KNN	RF
Training Time (ms)	0.00	15.61	7.00	174.00	0.00	21.20	7.82	169.31
Accuracy(%)	93.04	88.95	95.74	94.59	95.55	91.89	95.93	95.93
Precision(%)	95.96	87.78	93.94	95.56	95.56	88.88	94.56	94.79
Area Under Curve	0.99	0.95	0.95	0.94	1.00	0.96	0.88	0.910
Error (%)	2.12	8.10	4.44	2.12	2.33	4.25	3.12	2.32
Control Group Datasets								
Performance Metrics	NB	SVM	KNN	RF	NB	SVM	KNN	RF
Training Time (ms)	0.00	0.00	0.00	153.63	0.00	0.00	0.00	169.31
Accuracy (%)	92.30	93.94	84.84	93.93	84.61	84.84	74.35	92.30
Precision (%)	88.82	88.87	80.73	90.30	87.63	87.36	80.03	94.01
Area Under Curve	0.87	0.95	0.84	0.99	0.88	0.97	0.86	1.00
Error (%)	13.51	6.06	15.15	6.06	15.9	7.60	25.64	7.69

Figure 6. Algorithm Performance Metrics for Treatment Group: Cognitive Load



Article DOI: https://doi.org/10.37284/eajit.7.1.2227

Figure 7. Algorithm Performance Metrics for Control Group: Cognitive Load

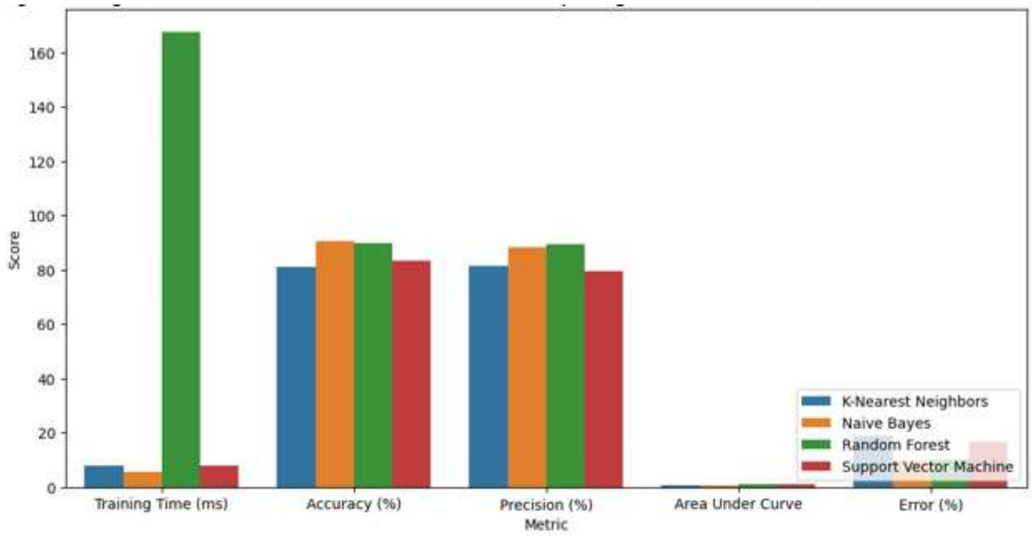
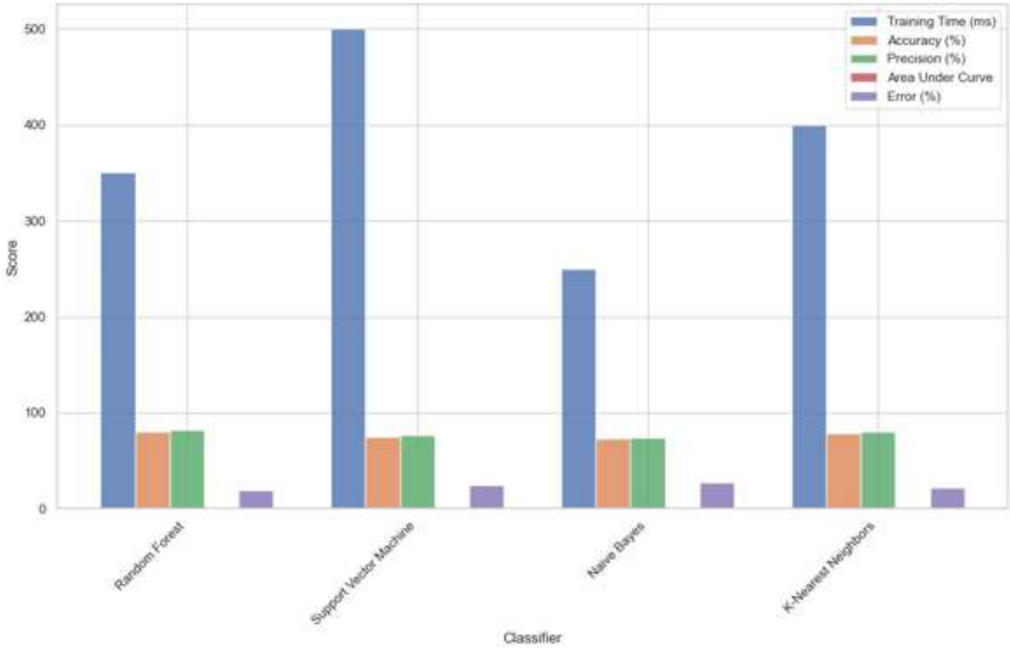


Figure.8 Algorithm Performance Metrics for Treatment Group



LIMITATIONS

This study had some limitations, including the experiment's short duration of four weeks may not capture long-term effects and trends in cognitive load and performance. Uncontrolled external factors, such as students' circumstances and varying study environments, could have influenced cognitive load and performance but were not accounted for in the analysis. While efforts were made to blind students to the interventions, complete blinding might not have been achievable, potentially influencing student behaviour and outcomes. Lastly, differences in training time and computational resources required by different machine learning models may affect the practicality of implementing these models in real-world educational settings. These limitations highlight areas for future research to enhance the robustness and applicability of the findings.

CONCLUSION

In conclusion, the study demonstrated that Naive Bayes (NB) and Support Vector Machine (SVM) models effectively predict both cognitive load and

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

performance outcomes in Moodle environments, with NB showing particularly strong precision and accuracy. SVM also performed well, though with slightly longer training times. Random Forest (RF), while accurate, required more computational resources. These findings suggest, leveraging NB and SVM for predictive analytics in e-learning platforms to enhance student engagement and academic performance. Based on these results, future studies should further explore the integration of cognitive load metrics into machine learning models to refine predictions and personalize interventions effectively in Moodle and similar LMS environments.

REFERENCES

Aeiad, E., & Meziane, F. (2019). An adaptable and personalised elearning system applied to computer. *Education and Information Technologies*, *78*, 674–681.

Boulesteix, A. L., & Strobl, C. (2009). Optimal classifier selection and negative bias in error rate estimation: An empirical study on high-dimensional prediction. *BMC Medical Research Methodology*, *9*, 1–14. https://doi.org/10.1186/1471-2288-9-85

Dobre, I. (2015). Learning Management Systems for Higher Education - An Overview of Available Options for Higher Education Organizations. *Procedia - Social and Behavioral Sciences*, *180*(November 2014), 313–320. https://doi.org/10.1016/j.sbspro.2015.02.122

Garkule, V., & Makarevičs, V. (2018). Moodle Environment and Its Use Within Formal and Informal Education At a Vocational School. *SOCIETY. INTEGRATION. EDUCATION. Proceedings of the International Scientific Conference*, *5*, 68– 80. https://doi.org/10.17770/sie2018vol1.3101

Garrison, D., Anderson, T., & Archer, W. (2000). Critical Inquiry in a text-based environment. *The Internet and Higher Education*, *2*(2), 87–105.

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *Internet and Higher Education*, *2*(2– 3), 87– 105. https://doi.org/10.1016/S1096-7516(00)00016-6

Geron, A. (2016). *Hand-on Machine Learning with Scikit-Learn,Keras and TensorFlow* (Issue 0).

Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U., & Sattar, M. U. (2020). Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Applied Sciences (Switzerland)*, *10*(11). https://doi.org/10.3390/app10113894

Ho, I. M. K., Cheong, K. Y., & Weldon, A. (2021). Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques. *PLoS ONE*, *16*(4 April), 1– 27. https://doi.org/10.1371/journal.pone.0249423

Hoque, E. M. (2016). Three Domains of Learning: Cognitive, Affective and Psychomotor. *The Journal of EFL Education and Research (JEFLER)*, *2*(2), 2520–5897. www.edrc-jefler.org

Hstie, T., Tibshiran, R., & Jerome, F. (2013). The Elements of Statistical Learning, Data Mining, Inference and Prediction. In *Springer*. https://doi.org/10.1109/SITIS.2013.106

Juhanie, T., & Paas, F. (2017). *Exploring Multidimensional Approaches to the Efficiency of Instructional Conditions Author (s): JUHANI E. TUOVINEN and FRED PAAS Source: Instructional Science, Vol. 32, No. 1 / 2, Special Issue: Advances in Cognitive Load Theory; Methodology an*. *32*(1), 133–152.

Kaensar, C., & Wongnin, W. (2023). Analysis and Prediction of Student Performance Based on Moodle Log Data using Machine Learning Techniques. *International Journal of*

Emerging Technologies in Learning, *18*(10), 184– 203. https://doi.org/10.3991/ijet.v18i10.35841

Kalyuga, S. (2012). Interactive distance education: A cognitive load perspective. *Journal of Computing in Higher Education*, *24*(3), 182– 208. https://doi.org/10.1007/S12528-012-9060-4

Kalyuga, S., & Liu, T.-C. (2015). Guest Editorial: Managing Cognitive Load in Technology-Based Learning Environments. *Journal of Educational Technology & Society*, *18*(4), 1–8. http://www.jstor.org/stable/jeductechsoci.18.4.1

Kante, M., Oboko, R., & Chepken, C. (2016). Factors affecting the use of ICT s on agricultural input information by farmers in developing countries. *AIMS Agriculture and Food*, *1*(3), 315– 329. https://doi.org/10.3934/agrfood.2016.3.315

Kolekar, S. V., Pai, R. M., & Manohara Pai, M. M. (2018). Adaptive User Interface for Moodle-based E-learning System using Learning Styles. *Procedia Computer Science*, *135*, 606– 615. https://doi.org/10.1016/j.procs.2018.08.226

Krishnan, R., Nair, S., Saamuel, B. S., Justin, S., Iwendi, C., Biamba, C., & Ibeke, E. (2022). Smart Analysis of Learners Performance Using Learning Analytics for Improving Academic Progression: A Case Study Model. *Sustainability (Switzerland)*, *14*(6), 1–13. https://doi.org/10.3390/su14063378

Lauren, G., & Vanessa, R. (2022). *Understanding Cognitive Load Theory for better online course design*. https://moodle.com/us/news/understanding-cognitive-load-theory-for-better-online-course-design/

Learning analytics indicators - MoodleDocs. (n.d.). https://docs.moodle.org/310/en/Learning_analytics_indicators

Młynarska, E., Greene, D., & Cunningham, P. (2016). *Indicators of Good Student Performance in Moodle Activity Data*. *January*. http://arxiv.org/abs/1601.02975

Oppong, S. O. (2023). Predicting Students' Performance Using Machine Learning Algorithms: A Review. *Asian Journal of Research in Computer Science*, *16*(3). https://doi.org/10.9734/ajrcos/2023/v16i3351

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). https://doi.org/10.22554/ijtel.v5i1.57

Rachel, V., Sudhamathy, G., & Parthasarathy, M. (2018). Analytics on Moodle Data Using R Package for Enhanced Learning Management. *International Journal of Applied Engineering Research*, *13*(22), 15580–15610. http://www.ripublication.com

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), 1–21. https://doi.org/10.1002/widm.1355

Schunk, D. H. (2012). Learning Theories,An Educational Perspective,Sixth Edition. In *Space Science Reviews* (Vol. 71, Issues 1–4).

Segura, M., Mello, J., & Hernández, A. (2022). Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role? *Mathematics*, *10*(18), 1–20. https://doi.org/10.3390/math10183359

Shayan, P., & van Zaanen, M. (2019). Predicting student performance from their behavior in learning management systems. *International Journal of Information and Education Technology*, *9*(5), 337–341. https://doi.org/10.18178/ijiet.2019.9.5.1223

Skulmowski, A., & Rey, G. D. (2017). Measuring cognitive load in embodied learning settings. In *Frontiers in Psychology* (Vol. 8, Issue AUG, p. 1191). Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2017.01191

Article DOI: https://doi.org/10.37284/eajit.7.1.2227

Song, J. H., & Thompson, L. (2011). *Ji Hoon Song, PhD*. *24*(3), 55– 76. https://doi.org/10.1002/piq

Sucharitha, G., Matta, A., & Dwarakamai, K. (2020). Theory and Implications of Information Processing. *Emotion and Information Processing*, *January 2021*. https://doi.org/10.1007/978-3-030-48849-9

Sun, J. C. Y., Liu, Y., Lin, X., & Hu, X. (2023). Temporal learning analytics to explore traces of self-regulated learning behaviors and their associations with learning performance, cognitive load, and student engagement in an asynchronous online course. *Frontiers in Psychology*, *13*(January). https://doi.org/10.3389/fpsyg.2022.1096337

Sweller, J., Ayres, P., & Kalyuga, S. (2011). The Goal-Free Effect. *Cognitive Load Theory*, 89–98. https://doi.org/10.1007/978-1-4419-8126-4_7

Theng, D., & Theng, M. (2020). *Machine Learning Algorithms for Predictive Analytics: A Review and New Perspectives*. *July*, 1–10. https://doi.org/10.37896/HTL26.06/1159

Tlili, A., Denden, M., Essalmi, F., Jemni, M., Chang, M., Kinshuk, & Chen, N. S. (2023). Automatic modeling learner's personality using learning analytics approach in an intelligent Moodle learning platform. *Interactive Learning Environments*, *31*(5), 2529– 2543. https://doi.org/10.1080/10494820.2019.1636084

van Mierlo, C. M., Jarodzka, H., Kirschner, F., & Kirschner, P. A. (2012a). Cognitive load theory in e-learning. In *Encyclopedia of Cyber Behavior* (Vol. 1, Issue December 2015). https://doi.org/10.4018/978-1-4666-0315-8.ch097

van Mierlo, C. M., Jarodzka, H., Kirschner, F., & Kirschner, P. A. (2012b). Cognitive load theory in e-learning. In *Encyclopedia of Cyber Behavior* (Vol. 1, Issue January). https://doi.org/10.4018/978-1-4666-0315-8.ch097

Wasylewicz, A. T. M., & Scheepers-Hoeks, A. M. J. W. (2018). Clinical decision support systems. In *Fundamentals of Clinical Data Science*. https://doi.org/10.1007/978-3-319-99713-1_11

Yen, A. C., Chen, I., Lai, S., Chuang, Y., Yen, C., Chen, I., Lai, S., & Chuang, Y. (2015). An Analytics-Based Approach to Managing Cognitive Load by Using Log Data of Learning Management Systems and Footprints of Social Media Linked references are available on JSTOR for this article : An A. *Educatinal Technology & Society*, *18*(4), 140– 158. file:///C:/Users/vdmerwer/Downloads/jeductechsoci.18.4.141.pdf

Yen, C.-H., Chen, I.-C., Lai, S.-C., & Chuang, Y.-R. (2015). An Analytics-Based Approach to Managing Cognitive Load by Using Log Data of Learning Management Systems and Footprints of Social Media. *Journal of Educational Technology & Society*, *18*(4), 141– 158. http://www.jstor.org/stable/jeductechsoci.18.4.141

Young, M. (2015). What is learning and why does it matter? *European Journal of Education*, *50*(1), 17– 20. https://doi.org/10.1111/ejed.12105

