African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

USING LMS LOG DATA TO IDENTIFY AT-RISK STUDENTS: A SYSTEMATIC REVIEW OF MACHINE LEARNING APPROACHES AND BIBLIOGRAPHIC ANALYSIS

Karim-Abdallah, B.[1], Weyori, B. A.[2], Mensah, P. K.[3]

[1]*Quality Assurance and Academic Planning Directorate, University of Energy and Natural Resources, Sunyani, Ghana.*

[2]*Department of Computer and Electrical Engineering, University of Energy and Natural Resources Sunyani, Ghana.*

[1,3]*Department of Computer Science and Informatics, University of Energy and Natural Resources Sunyani, Ghana.*

[1]*bright.karim-abdallah@uenr.edu.gh*

ABSTRACT

Purpose: This study evaluates the effectiveness of machine learning algorithms in predicting student dropout using Learning Management System (LMS) log-in data.

Design/ Methodology/ Approach: The study used a systematic literature review and bibliographic analysis. The search encompassed papers from the Scopus database up to October 2024. Initially, 100 articles were identified. After applying exclusion criteria, including removing editorials, letters, comments, and conference papers, 61 studies were chosen for the final review. Performance criteria such as accuracy, precision, recall, and f1-score were employed to assess these studies.

Research Limitation: Several limitations were acknowledged, including potential publication bias due to the inclusion of only peer-reviewed articles, variability in educational contexts and LMS platforms, and heterogeneity in machine learning methods and performance metrics.

Findings: Random Forest emerged as the most commonly used machine learning algorithm for identifying at-risk students, followed by Convolutional Neural Networks. In the analysed research, Random Forest outperformed all other algorithms, achieving a 99% accuracy rate in predicting at-risk students. Students' assessment scores emerged as the most significant feature in the model performances, followed by students' participation in a session.

Practical Implication: It is noted that most researchers do not report on significant features/variables or the contribution of features to the model's performances.

Social Implication: These significant features or variables are essential for institutions employing a blended learning approach, as they provide insights into where to allocate limited resources most effectively.

Originality/Value: This study contributed to the pool of knowledge on how Machine learning techniques have been employed with Learning Management System log-in data to predict student dropout.

Keywords: *At-Risk Students. drop-out. e-learning. learning management system. log data.*

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

INTRODUCTION

Educational institutions are increasingly compelled to enhance their student's academic performances, with graduate employability as a key metric for assessing the effectiveness of their study programmes. These institutions provide future learning opportunities and contribute to global efforts to address the rapid evolution of social, cultural, economic, and environmental sustainability challenges in the 21st century (Mertova & Nair, 2011). This mandate aligns with the achievement of Sustainable Development Goal 4 (SDG 4), which advocates for inclusive, equitable, and quality education and promotes lifelong learning opportunities for all. The realisation of this goal hinges on universities producing graduates adequately equipped to navigate the uncertain demands of the future.

In recent decades, substantial research has been conducted on predicting student success (Bañeres et al., 2023). The internet now serves as the village square in today's global village, facilitating almost every activity (Karim-Abdallah & Harris, 2022), including education. Online learning, also known as e-learning, has greatly benefited from the Internet's extensive use (Ahmed, 2024).

Tertiary institutions and governments have expanded access to accommodate the rising demand for post-secondary education. However, many students seeking further education are from the working class, leading to exponential growth in student numbers annually. Consequently, Distance Blended and Online Learning (DBOL) has been introduced (Karim-Abdallah et al., 2025a). In Sub-Saharan Africa, increased accessibility options include distance and transnational education (Prahani et al., 2022).

The emergence of COVID-19 saw the use of online platforms like Google Classroom, Zoom, Google Meet, and Moodle to ensure the continuation of education. Schools also adopted LMS to facilitate student participation in online activities. LMSs track student activities, from logging in to taking quizzes, submitting assignments, and completing exams. Researchers have utilised these logs to forecast students at risk of early dropout (Arizmendi et al., 2023; Bañeres et al., 2023; Figueroa-Canas & Sancho-Vinuesa, 2020; Porras et al., 2023; Tan & Shao, 2015; Yousafzai et al., 2021).

To predict student success in courses or programmes, machine learning approaches have been used in numerous studies (Ahajjam et al., 2022; Ahmed, 2024; Bañeres et al., 2023; Chung & Lee, 2019; Dasi & Kanakala, 2022; Fu et al., 2021; Porras et al., 2023; Tan & Shao, 2015; Yousafzai et al., 2021; Zhen et al., 2023; Zhou & Xu, 2020). This review examines various investigations to identify the model that most accurately predicts data from online studies. Machine learning frequently aims to find a single model that most closely predicts the desired outcome (Karim-Abdallah et al., 2025b). Consequently, several methods are investigated, each providing distinct outcomes for use. The "black box" aspect of machine learning techniques has drawn criticism since it restricts the interpretability of models to forecast student dropout. Therefore, it is crucial to carry out an extensive, methodical evaluation of the use of machine learning to forecast student dropout. This research aims to determine which ML algorithms have the most predictive power when combined with LMS Log Data.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

A comprehensive search has been done to locate systematic research reviews on student dropouts. To our knowledge, no extant research explicitly addresses our specific objective. This study aims to bridge this gap in the literature and identify the most effective machine-learning strategy for predicting student dropout using LMS log data.

RELATED WORKS

This section is devoted to an overview of extant research and reviews focused on LMS log data and machine learning algorithms.

Arizmendi et al. (2023) provide a comprehensive review of the use of LMS digital data to predict student success in higher education. The study focuses on the growing trend of using digital traces from LMS platforms, such as time-stamped logs of student behaviour, to forecast course results and advise interventions. The authors examine multiple machine learning algorithms used to model student success, comparing performance metrics, including prediction accuracy, specificity, and F-scores among models. Furthermore, the study investigates ethical implications, including the inclusion of demographic factors, which raises concerns about justice and equity in predictive models. The authors emphasise the importance of feature creation and offer improvements to feature generation to improve prediction power and produce actionable insights. Furthermore, the report suggests that future research should focus on enhancing model interpretability, removing algorithmic biases, and using advanced approaches such as regularisation to improve predictions while maintaining fairness. The study could not identify the most often utilised machine learning techniques in the LMS area.

Wang and Mousavi (2023) did a study that included a comprehensive evaluation and meta-analysis of the association between log variables obtained from online learning platforms and student academic progress. The authors examined 88 empirical papers published between 2010 and 2021 to determine critical log variables strongly affecting student performance in online learning environments. The study divided log variables into two types: basic and complicated. Basic variables included frequency and duration spent on tasks (e.g., login time, page visits), whereas complex variables recorded more in-depth patterns. Login time, frequency of page views, and time-on-task all correlated positively with academic success. In contrast, irregular study intervals and late submissions were related to lower achievement rates. Login time, frequency of page views, and time-on-task all correlated positively with academic success. In contrast, irregular study intervals and late submissions were related to lower achievement rates. The study found that learning modality (totally online vs. hybrid), course theme (e.g., STEM, business), and sample size all impacted log variable predictive power. The research emphasises the necessity of analysing student behaviour using log data to improve predictive modelling for academic performance. The findings can help instructors and institutions build more effective learning interventions and support systems using data-driven insights. Furthermore, the work emphasises the

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

necessity for future research to investigate complex log variables to make more nuanced predictions. The study did not examine the most effective machine learning methods for predicting LMS log data.

Bervell and Umar (2017) conducted a comprehensive research analysis on LMS acceptance and use in Sub-Saharan Africa (SSA) from 2007 to 2017. The study aimed to bring together disparate findings to understand better research patterns, important variables, and problems in LMS adoption in higher education institutions in Southern Africa. The authors discovered that the Technology Acceptance Model (TAM) was the most commonly utilised framework in LMS studies, accounting for 58.1% of the reviewed articles. The Unified Theory of Acceptance and Use of Technology (UTAUT) was the second most popular model. The review revealed that students were the primary topic in most articles (58.1%), with less research on instructors (19.4%). The most common research approach was quantitative (80.6%), with questionnaires as the primary data-gathering method. The study found that attitude, perceived usefulness, and performance expectancy were the most important predictors of LMS adoption. Social influence plays an important effect in setting usage intentions. Major impediments to LMS deployment were inadequate ICT infrastructure, a lack of skills/training, and system-related concerns. Furthermore, essential variables for successful implementation included leadership, management support, and explicit LMS use policies. The review underscores the need for further research focusing on instructors' adoption behaviour and advocates for using advanced models such as TAM2, TAM3, and UTAUT in future studies. It also recommends a shift toward mixed-methods research to provide deeper insights into the complexities of LMS adoption. This work is highly relevant for understanding the broader context of LMS implementation challenges in SSA and informs future studies on effective adoption strategies. The authors concluded that statistical analyses often employed regression models to assess LMS adoption determinants.

METHODS

The current study's systematic review methodology followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) criteria, and VosViewer was used for bibliographic analysis. Five (5) essential phases comprised the systematic review:

1. Formulation of Research Question (RQ): The study's overarching research question was developed.
2. Development of Research Protocol: A research protocol detailing the systematic review's design was constructed. This protocol covered the selection criteria for the studies, the databases used, search tactics, and techniques for extracting and analyzing data. The studies' inclusion and exclusion criteria were also established.
3. Literature Search: A credible database was used to investigate scientific literature thoroughly. Specific search phrases and inclusion and exclusion criteria were utilised to choose relevant studies. Papers that satisfied the inclusion requirements were selected. The chosen studies were

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

carefully reviewed to ensure they satisfied all requirements. Important information was taken out of these investigations and organised into a database.

4. Data Synthesis and Interpretation: Possible limitations were highlighted once the results and findings were interpreted.
5. Reporting: The systematic review was written up in a complete report that included a discussion of review outcomes

Identify the source

This systematic literature review (SLR) retrieved the primary studies using the Scopus database. The Scopus (https://www.scopus.com/) database provides comprehensive, well-vetted information on various subjects and is the largest database of scholarly articles.

Table 1: Search Strategy

Database	Search Terms	Coverage Period
Scopus	(TITLE-ABS-KEY ("Machine learning" OR "Deep Learning") AND TITLE-ABS-KEY ("student dropout prediction " OR "Student desertion" OR "student dropout" OR "Academic performance") AND TITLE-ABS-KEY ("LMS log data" OR "e-learning")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (PUBSTAGE , "final"))	2014 to October,2024

Criteria for Inclusion and Exclusion

Which research or articles are included in the review, and those excluded are determined by the standards and guidelines used in this scientific investigation. The goals and inquiries of the research set these standards.

Inclusion Criteria:

i. Research using LMS log data and ML algorithms to identify at-risk students or predict dropout.
ii. Papers that have been printed in peer-reviewed journals.
iii. Research presenting performance measures, including specificity, recall, sensitivity, accuracy, and precision.
iv. Studies available in English.

Exclusion Criteria:

i. Conference papers, editorials, letters, and comments.
ii. Studies lacking sufficient detail on the machine learning methods used or their performance metrics.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

iii. Studies focused on non-academic dropout factors or without a clear connection to LMS log data.
iv. Studies with no full text available

Information was obtained from each research using a standardized form, capturing details such as Publication Details: Author(s), year, and source. Study Context: Educational setting, sample size, and LMS type. Machine Learning Methods: Algorithms used, features, Performance Metrics: Accuracy, precision, sensitivity, recall, specificity, etc. Additional Insights: Key findings or observations related to the effectiveness of the machine learning models.

A narrative synthesis approach was employed to summarise the findings. The analysis focused on identifying commonly used machine learning algorithms, comparing model performance, highlighting best practices, and discussing model interpretability and practical implications. Performance metrics were aggregated where possible to provide an overview of algorithm effectiveness. Trends and patterns were identified and discussed to draw meaningful conclusions.

Several limitations were acknowledged, including potential publication bias due to the inclusion of only peer-reviewed articles, variability in educational contexts and LMS platforms, and heterogeneity in machine learning methods and performance metrics. Despite these limitations, the review aims to provide a thorough overview of the present state of study on predicting at-risk students using LMS log data and machine learning approaches.

Procedure for Selecting Samples

The sample was reduced to only contain items relevant to the study's goal once the inclusion and exclusion criteria were put into practice. According to a flow graphic with the study, one hundred items were initially located throughout the database. Forty-six articles were eliminated due to duplicates, records marked as ineligible by automation tools and application of the defined criteria. There were 61 publications total that were included in the analysis after more exclusions were made for different reasons (Figure 1).

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

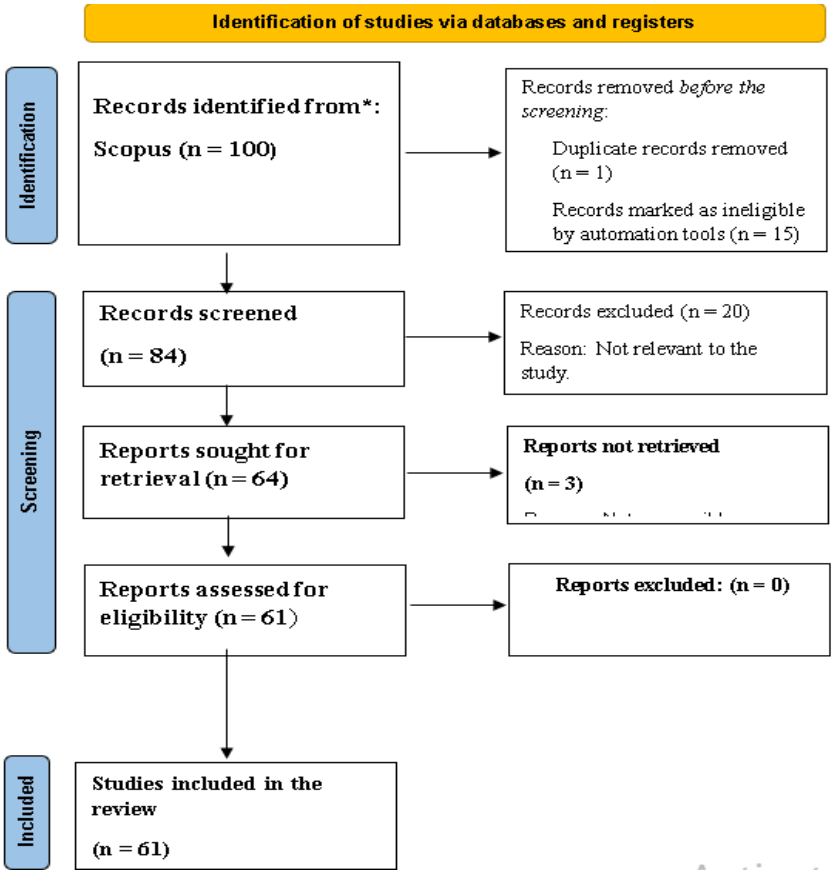

Figure 1 displays a PRISMA flowchart illustrating the process of searching and selecting references for the systematic review.

FINDINGS AND DISCUSSION

Table 3 shows the pertinent features of the studies included in the systematic review. The attributes considered are author, year, title, nation, sample size, number of variables, Machine Learning methods used, top-performing algorithm, performance metrics, and outcomes (accuracy, precision, recall, F1 score).

RQ1: What are the publication trends of machine learning methods in predicting at-risk students using LMS log data?

The findings presented in Figure 2 illustrate the allocation of primary papers based on their continent of origin and respective countries. In this study, three papers, making up 4.92% of the total, originated from North America, specifically from the US (2) and Canada (1). Europe had the second highest number of papers, with a total of fourteen originating from Spain (5), Hungary (2), Slovakia (2), France (1), Greece

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

(1), Finland (1), Italy (1) and Romania (1), accounting for 22.95% of the primary papers. Asia also contributed thirty-three papers, recording the highest number of publications during the year under review, 54.1 % of the total, with representation from China (12), Pakistan (4), Hong Kong (1), Saudi Arabia (4), Iraq (1), Indonesia (1), Jordan (1), Japan (1) and India (8). Four papers came from South America, one from Brazil and three from Ecuador, representing 6.56%. Africa had six papers (9.84), with Morroco contributing three papers, Egypt, Ethiopia and Nigeria recording one paper each.

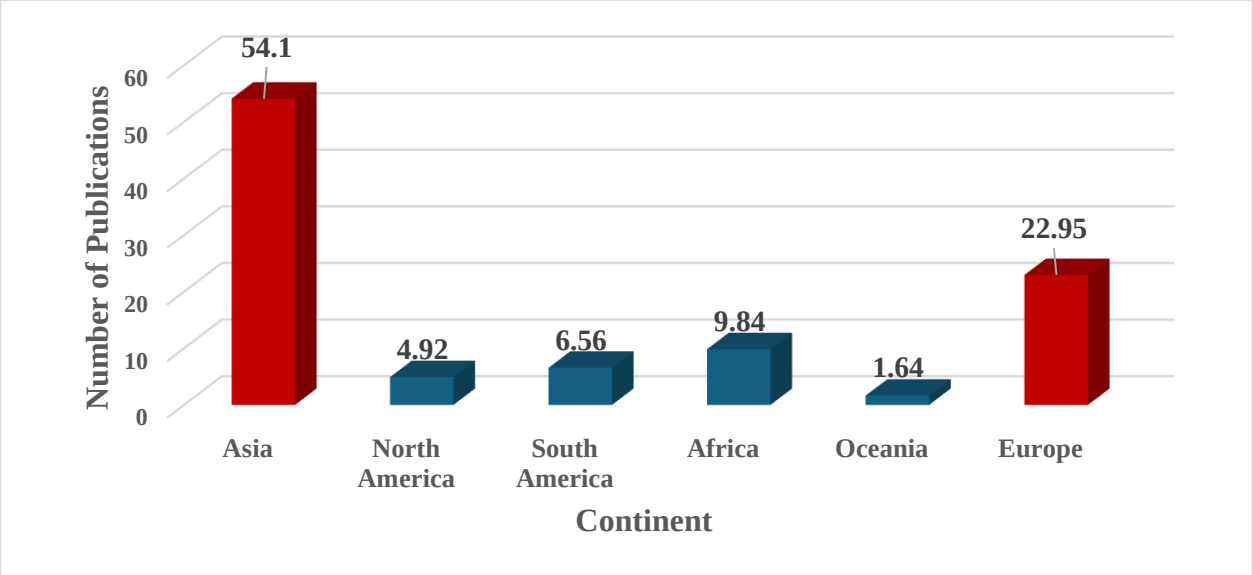

Figure 2: Percentage Distribution of Papers Across Continents

One paper, representing 1.64% of the total, originated from Oceania (Australia). The data indicates that Europe and Asia are the leading regions in research on predicting student dropout using LMS log data and Machine Learning.

Figure 3 illustrates the number of articles published within the selected period. The number of publications from 2014 to 2019 hovered around 1 (1.64%). The line graph shows an increase in publications from 1 in 2019 to 9 (14.75%) in 2020. The number of publications increased to 11(18.03) in 2021. This number decreased to 9 (14.75%) publications in 2022, subsequently rising to 18 (29.51%) in 2023. From 2023 to October 2024, it decreased to 11(18.03%) publications. This number may increase by the end of 2024.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

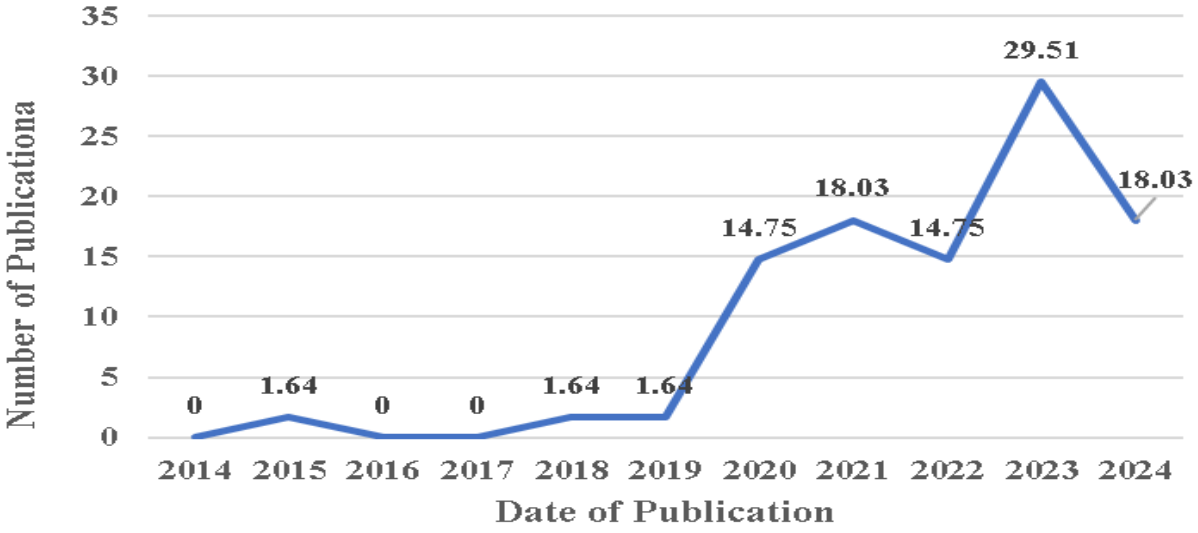

Figure 3: Publication year of the primary papers (2016 to October 2024)

Figure 4 shows the journals publishing machine learning techniques with LMS data. The IEEE Access contributed 6 (9.84%) papers, followed by the International Journal of Emerging Technologies in Education with 5 (8.20%) and Sustainability(Switzerland) with four papers (6.56%). Computers in Human Behaviour Journal and Education and Information Technology Journals had two papers (3.28%) each. Advances in Educational Physiology, Dyna(Spain) and forty remaining Journals contributed a paper each.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

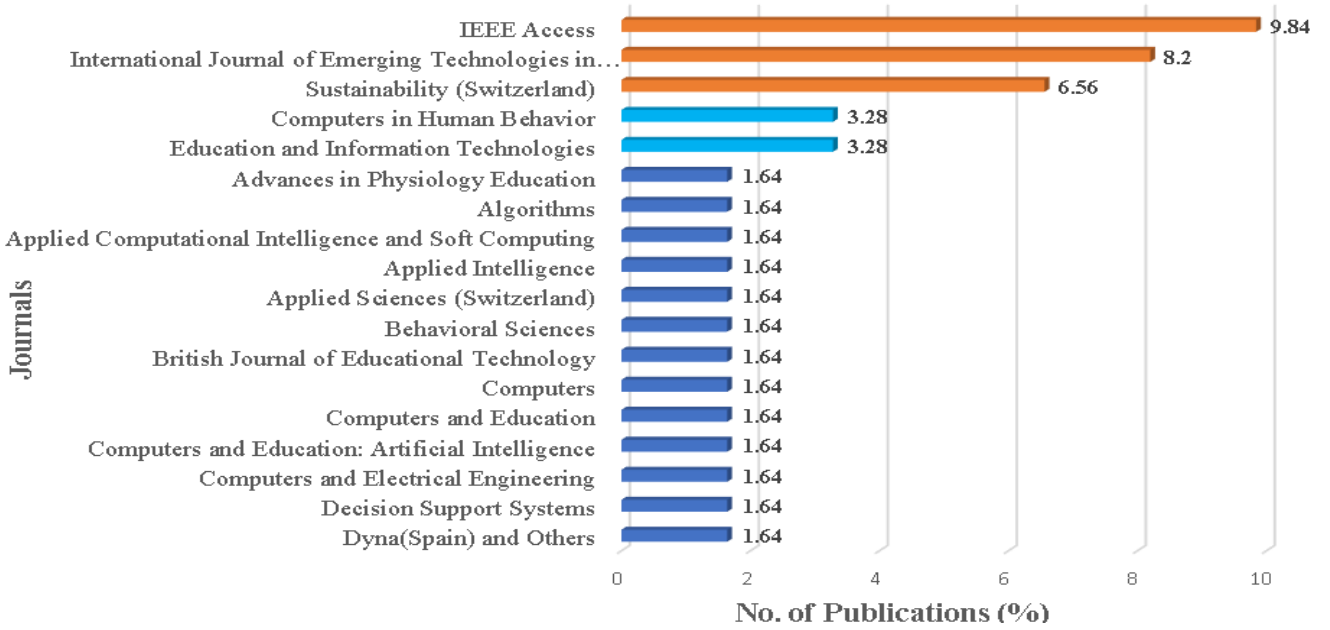

Figure 4: Publication outlets from the primary papers.

RQ2: What ML methods are used in predicting or classifying at-risk students using LMS log data?
Figure 5 illustrates the prevalent machine learning algorithms that outperformed other methods employed to analyse LMS log data. In total, 22 ML methods or algorithms have been employed using LMS log-in data to predict student dropout. Among these, Random Forest is the most commonly used algorithm, appearing in 9 out of 61 primary study papers, constituting 14.75%. Convolutional Neural Network and Neural Network each appear in 7 papers, representing 11.48% of the total, highlighting their popularity and effectiveness in predicting student dropout. Long Short-Term Memory is featured in 5 papers, accounting for 8.20% of the total. Decision Tree, Gradient Boosting, and Multi-layered Perceptron appear in 4 papers, constituting 6.56%. Support Vector Machine is used in 3 papers, making up 4.92% of the total, and is known for its effectiveness in classification tasks. Voting Algorithms also appear in 3 papers, representing 4.92% of the total, and are valued for their powerful modelling capabilities, especially in complex data scenarios. AdaBoost, Naïve Bayes, and Logistic Regression each appear in 2 papers. Additionally, as depicted in Table 2, eight different algorithms each appear in 1 paper.

Table 4 in the appendix contains the performance metrics of the primary study papers. The accuracy of the selected algorithms varies across the 61 studies, ranging from 77.78% to 100%, with the highest accuracy achieved by K-Nearest Neighbour and Long Short-Term Memory in two studies, PS1 and PS43, respectively, each reporting 100%. The lowest accuracy, 77.78%, was reported for Artificial Neural Networks in PS26. Random Forest, the most popular algorithm, shows strong performance with accuracy rates of up to 99%, as seen in PS5, PS6, PS12, PS13, PS28, PS30, PS32, PS40, etc. Neural Networks

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

also show strong performance with accuracy rates of up to 96%, as seen in PS26, PS49, PS51, PS55 etc. Convolutional Neural Networks and Long Short-Term Memory frequently appear among the chosen algorithms, with accuracy rates typically around 85% to 99.98%. These results underscore Random Forest as a consistently high-performing algorithm in predicting student success and dropout.

Two main issues arise in determining the optimal method for machine learning. First, finding the ideal algorithm is difficult because many candidate systems satisfy particular requirements (Dasi & Kanakala, 2022.; Zhou & Xu, 2020). This problem is made more difficult when there are few training data, and the learning algorithm shows a tendency for different local optima. Moreover, discarding less successful models carries the danger of losing potentially important information (Zhou & Xu, 2020).

The Random Forest algorithm has been widely recognised for its capacity to tackle the challenges of local optima and overfitting and its potential to leverage diversity and independence among trees (Kamal & Ahuja, 2019; Saluja et al., 2023). It has demonstrated superior performance in predicting student dropout, making it a suitable choice for this application (Saluja et al., 2023). The algorithm's effectiveness in feature subset selection, classification, and regression has also been highlighted (Saluja et al., 2023).Its reliability and effectiveness in various fields, including student performance analysis, financial crisis prediction, and disease risk prediction, have been emphasised (Zhou & Xu, 2020). However, the lack of interpretability in Random Forest models has been noted as a limitation (Adejo & Connolly, 2018). However, the method has been suggested as a substitute for regression in institutional research for its simplicity, cheap computational requirements, accurate predictions, adaptability, and interpretability (Alsulami et al., 2023).

When forecasting children in danger, the Random Forest algorithm performs better than other machine learning algorithms. It is a good choice because it can manage local optima and overfitting problems and use the independence and diversity among the trees. In machine learning, the ensemble approach, of which Random Forest is an example, is beneficial for combining multiple models to improve system robustness and predictive ability, lessen the possibility of choosing the wrong premise, and expand the premise space to estimate the objective function more accurately.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

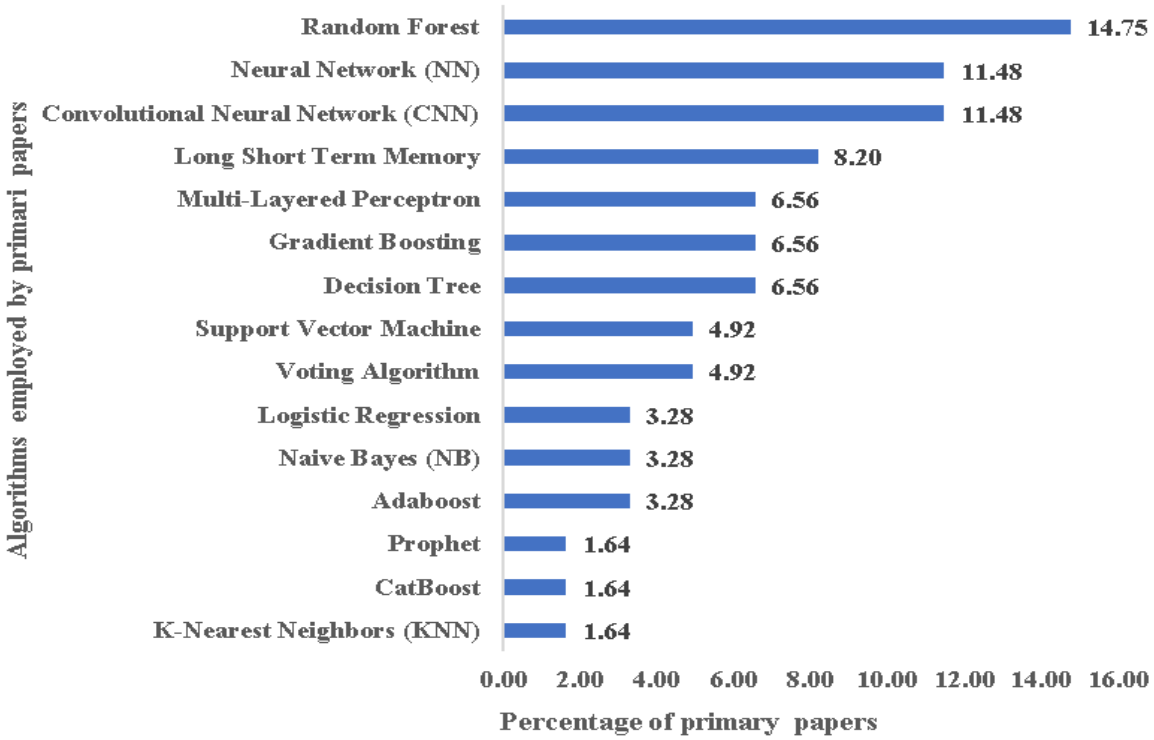

Figure 5: Percentage Distribution of Papers Across Widely Used ML Algorithms with LMS log data

Table 2: Widely used ML algorithms with LMS log data

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

S/N	Algorithm/Model	Primary Study Papers (PS)	Count	Percentage(%)
1	Random Forest	PS5, PS6, PS12, PS13, PS28, PS30, PS32, PS34, PS40	9	14.75
2	Convolutional Neural Network (CNN)	PS2, PS9, PS11, PS21, PS36, PS61,PS46	7	11.48
3	Neural Network (NN)	PS8, PS26, PS27, PS49, PS51,PS55,PS60	7	11.48
4	Long Short Term Memory	PS10,PS23, PS34,PS43,PS58	5	8.20
5	Decision Tree	PS14, PS17, PS22,PS53	4	6.56
6	Gradient Boosting	PS4, PS20, PS24, PS42	4	6.56
7	Multi-Layered Perceptron	PS39, PS41, PS52,PS57	4	6.56
8	Voting Algorithm	PS3, PS15, PS31	3	4.92
9	Support Vector Machine	PS7,PS16,PS59	3	4.92
10	Adaboost	PS35, PS44	2	3.28
11	Naive Bayes (NB)	PS34, PS48	2	3.28
12	Logistic Regression	PS18, PS19	2	3.28
13	K-Nearest Neighbors (KNN)	PS1	1	1.64
14	CatBoost	PS54	1	1.64
15	Prophet	PS 56	1	1.64
16	Attention-based Temporal Graph Network(AppTGN)	PS29	1	1.64
17	Gated Recurrent Unit-Autoencoder(GRU-AE)	PS47	1	1.64
18	Hierarchical Deep Learning(HDL)	PS37	1	1.64
19	Multi-Stage Spatial-Temporal Recurrent Attention (MSSRA)	PS33	1	1.64
20	K-Star	PS38	1	1.64
21	Repurposed transfer	PS45	1	1.64
22	Ensemble Stacking	PS50	1	1.64

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

RQ3: What are the significant features or attributes in these datasets?

The dataset sizes across the 61 selected articles in this study show significant variation. The total number of students is 42,074,273, with an average of approximately 725,419 students per study. The smallest dataset includes 20 students, while the largest comprises 40,000,000 students. Regarding the number of attributes used in these studies, there are 1,173 attributes across all studies, with an average of 23 attributes per study. The number of attributes ranges from a minimum of 3 to a maximum of 122, indicating some variation in the number of attributes used. This variation in dataset sizes and number of attributes highlights the diverse approaches and scales of LMS data used in predicting at-risk students using ML techniques.

Figure 6 and Table 3 depict the study's most significant attributes or features. Of the 61 primary studies in this review, only 27 papers, representing 44.26%, reported the most significant attributes or features in their models. This indicates that most machine learning models in the reviewed studies do not utilise feature importance to determine the contribution of each attribute within the model. Eight attributes or features, namely students' assessment scores, students' participation, number of times a student accesses a course, time spent studying, hints given to students, lack of knowledge of technical matters, video clips, and course content, are the most significant features in these 27 papers.

As illustrated in Figure 6, the most significant feature reported in the reviewed studies is students' assessment scores, highlighted by 37.04% of the primary study papers. This is followed by students' participation and the frequency of course access, with 22.22% and 18.52% of the studies reporting these features, respectively. These significant features or variables hold substantial importance for institutions employing a blended learning approach, as they provide insights into where to allocate limited resources most effectively.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

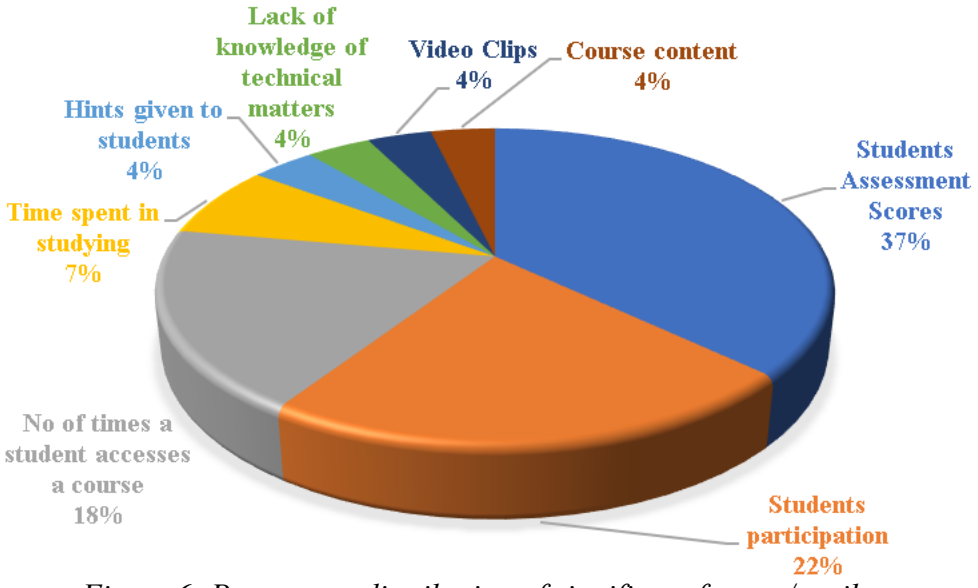

Figure 6: Percentage distribution of significant feature/attribute

Table 3: Significant attribute/feature in the dataset

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

S/N	Significant Variable/Feature	Primary Study Papers (PS)	Count	Percentage(%)
1	Students Assessment Scores	PS3,PS5,PS6,PS12,PS27, PS30,PS44,PS52,PS54,PS58	10	37.04
7	Students participation	PS18,PS40,PS41,PS56,PS57, PS59	6	22.22
6	No of times a student accesses a course	PS19,PS24,PS32,PS39,PS50	5	18.52
4	Time spent in studying	PS13,PS55	2	7.41
2	Hints given to students	PS4	1	3.70
3	Lack of knowledge of technical matters	PS11	1	3.70
5	Video Clips	PS25	1	3.70
8	Course content	PS49	1	3.70

RQ4: What gaps exist in the current literature?
Although online education or e-learning has been established for some time, the widespread lockdowns in most countries due to COVID-19 in 2020 necessitated the adoption of LMS by institutions for effective teaching and learning. This shift has led to more publications focusing on applying machine learning methods to predict student dropout.

Between 2000 and 2023, most ML techniques were employed to predict student dropout using LMS data. The high cost and weak internet connectivity in many regions of the Global South have significantly impacted institutions' ability to effectively utilize LMS for teaching and learning.

Challenges and issues associated with LMS continue to evolve, potentially impacting student performance. Today, the use of AI has a significant impact across various domains. As a subset of AI, ML techniques facilitate decision-making by learning from past events and forecasting future outcomes. According to Table 3, most researchers do not report on the significant variables or the contribution of each attribute to model performance. This information is crucial for the management of educational institutions to make informed decisions on targeted interventions to minimise the dropout rate. Furthermore, researchers may employ statistical models that account for the contribution of each variable to the model. Additionally, based on some findings from our primary research, effectively training students on using LMS is crucial for minimising dropout rates.

Analysis of Graphic Maps

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

This section uses bibliographic data to analyse scientific graphical maps using VOSviewer software (van Eck & Waltman, 2019). This research's two main mapping strategies are bibliographic coupling and co-citation. A technique for figuring out how similar two documents are thematically is co-citation coupling. This link is formed when a third document cites both texts, suggesting a shared subject matter. For instance, even in the absence of direct citations between publications A and B, a relationship between them can be inferred if paper C cites both papers A and B. The quantity of publications that co-cite A and B positively correlates with the intensity of this association.

Co-citation of Authors

The author co-citation analysis aims to reveal the structure and relationships among authors most frequently cited together. Figure 7 presents a co-citation network comprising four distinct clusters (blue, green, red and yellow colours) formed by co-citation links among 42 authors. The threshold for inclusion was set at a minimum of 10 citations, with 42 authors meeting this criterion out of a total of 5,673 authors cited in the references of primary study papers. Cristobal Romero (romero c.) emerges as the most cited author in the domain of LMS, with 29 citations and a total link strength of 416, as indicated by the prominent node in the first cluster (blue). Another highly cited author is Sebastian Ventura (ventura s.) in Cluster 1(blue) with 28 citations. In Cluster 2 (green), zdrahal z. is cited 20 times with a total link strength of 431, followed by Wu Zhang (zhang w.) in Cluster 3 with 19 citations and a total link strength of 278.

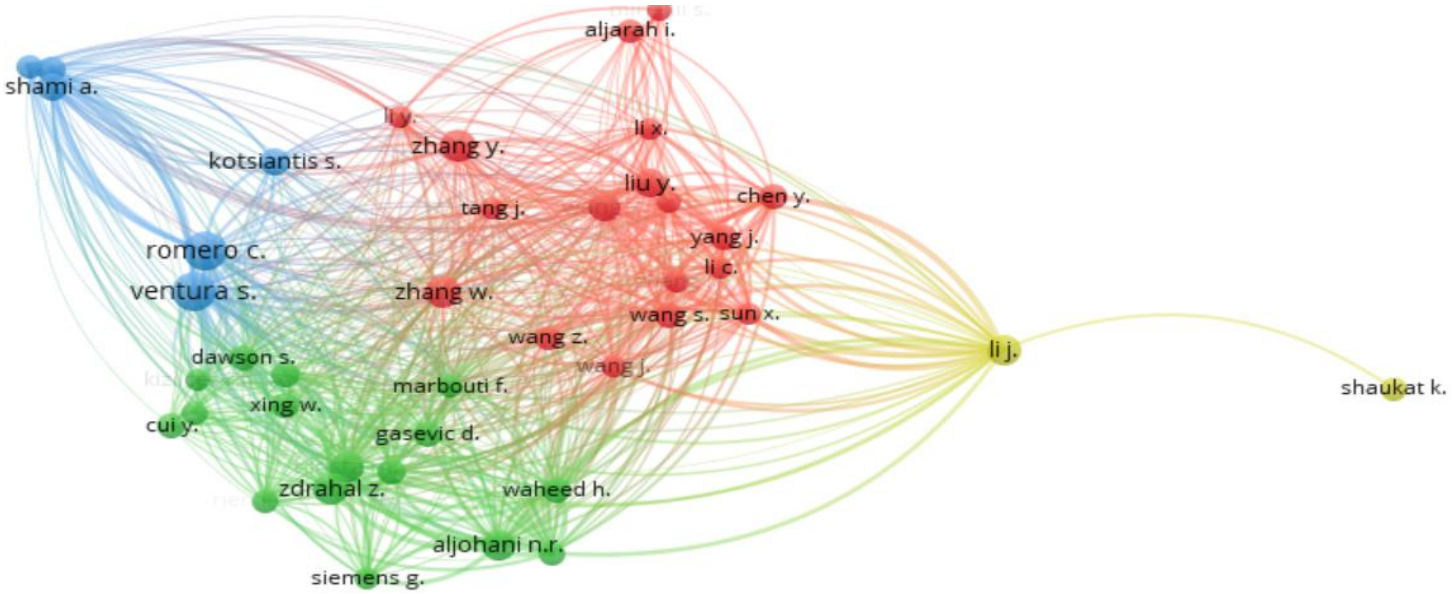

Figure 7: Co-citation of authors. The colours represent clusters

Co-Citation of Cited References

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

Figure 8 displays a co-citation network with two distinct clusters (red and green) formed by co-citation links among five references. The inclusion threshold was set at a minimum of three citations, resulting in seven references meeting this criterion out of 2,678 cited references in the primary study papers. Out of the 7 qualified references, 2 of them are connected. This analysis yielded 6 links. Notably, Waheed et al. (2020) paper in Cluster 1(red), *"Predicting academic performance of students from VLE big data using deep learning models?"* published in Computers in Human Behavior, is connected with all 4 papers in the map cited 3 times with 4 total link string as indicated with 4 lines in the map. In Cluster 2(green), Palmer's 2013 paper(Palmer, n.d.)"Modelling Engineering Student Academic Performance using Academic Analytics," published in the International Journal of Engineering Education, had three citations and a total link strength of 2. Figure 8 clearly shows that five papers are connected to our research topic.

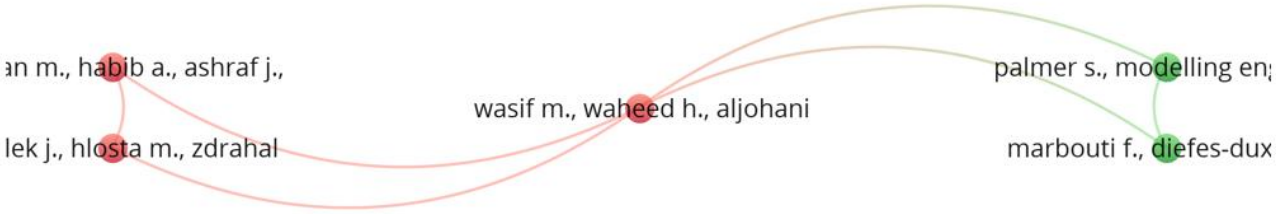

Figure 8: Co-citation of Cited References

Bibliographic Coupling of Countries

Figure 9 illustrates the bibliographic connections among significant countries, organised into two distinct clusters, each denoted by a different colour. These clusters are formed based on interconnections among five countries. The inclusion threshold was set at a maximum of twenty-five documents and a minimum of five documents per country, resulting in five countries meeting this criterion out of the 39 countries represented by corresponding authors in the references of the primary study papers. This analysis yielded 10 links with a total link strength of 273. Notably, China stands out with 14 documents and 396 citations, achieving a total link strength of 135. India follows with 8 documents and 108 citations, a total link strength of 15. The map reveals intricate bibliographic interconnections among these 5 countries, notably showing that China, India, Pakistan and Spain are bibliographically linked to Saudi Arabia in the context of our research.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

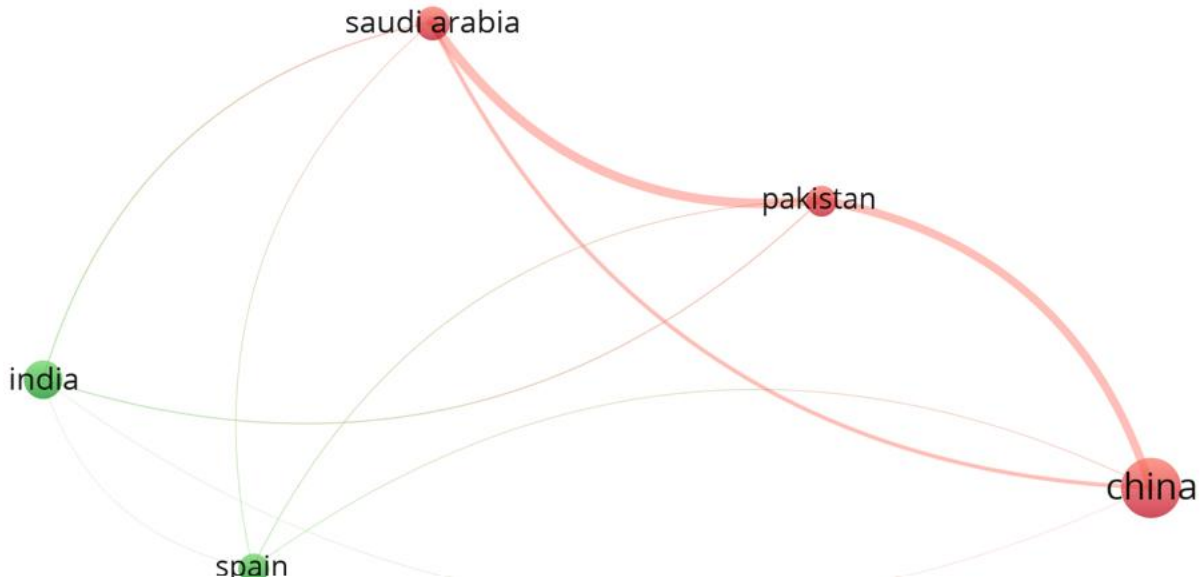

Figure 9: Bibliographic coupling by countries.

Co-occurrence of Author Keywords

Figure 10 presents a co-occurrence network of authors' keywords, organised into three distinct clusters represented by different colours. The inclusion threshold was set at a minimum of five occurrences, resulting in nine keywords meeting this criterion out of 197 from the primary study papers. This analysis yielded 32 links with a cumulative link strength of 53. The term "machine learning" has the most significant node, indicating its frequent use as a keyword, it appears 25 times. The size of each node corresponds to the frequency of the keyword's usage. These keywords can also be utilised to determine the types of papers included in the review.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

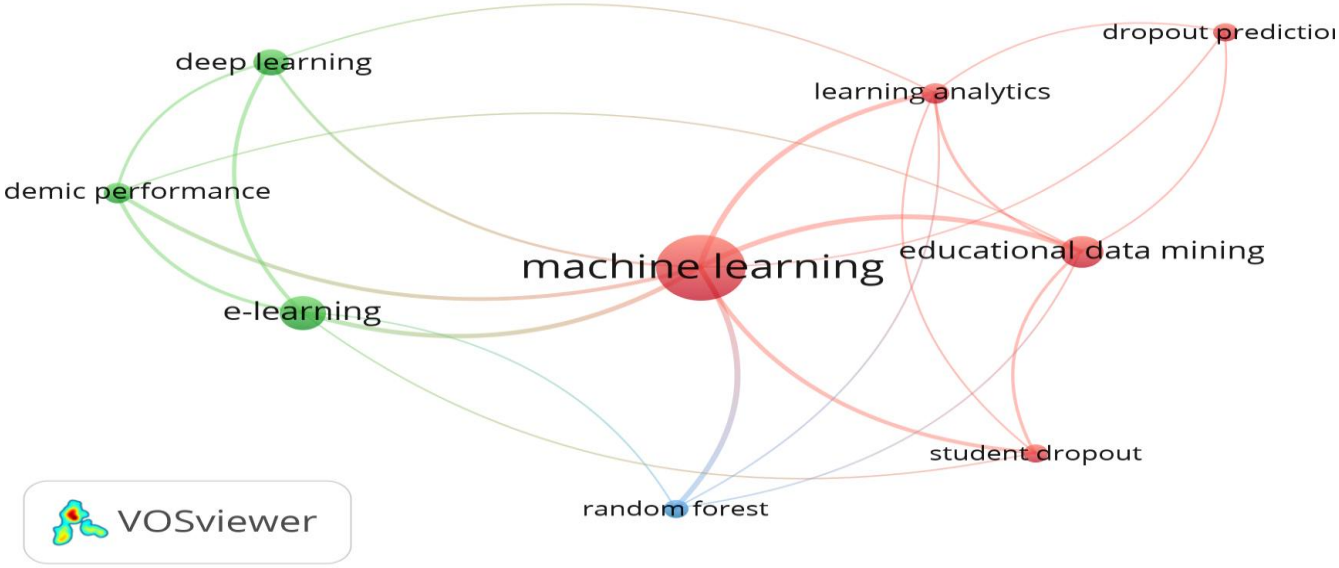

Figure 10: Co-occurrence of author's keyword

CONCLUSION

The review on predicting at-risk students using LMS log data highlights the significant potential of ML algorithms in identifying students at risk of dropping out. The analysis of 61 studies from various countries shows that ML techniques, especially supervised learning algorithms, are widely used and effective in educational contexts. Random Forest, in particular, stands out as the most frequently employed and highest-performing algorithm, achieving accuracies as high as 99% in some selected articles in this study. This consistent performance across different datasets and educational settings underscores the robustness and reliability of Random Forest in dropout prediction tasks.

The review also reveals that ML algorithms like Convolutional Neural Networks, Decision Trees, Long Short-Term Memory and Neural Networks are commonly used alongside Random Forests. These algorithms also demonstrate strong performance metrics, although slightly lower than Random Forest. The studies included in the review collectively involved a diverse set of variables and substantial sample sizes, averaging around 725,419 students per study. This extensive data highlights the generalizability of the findings and the applicability of ML algorithms in various educational settings, ranging from higher education institutions in China and Europe to emerging educational contexts in Asia and Africa.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

Ultimately, the application of ML algorithms to predict at-risk students offers a valuable tool for educational institutions aiming to improve student retention and success. By leveraging LMS log data, institutions can proactively identify and support students who might be struggling, thereby enhancing the overall educational experience and outcomes. The findings of this review suggest that continued development and implementation of ML-based predictive models hold promise for addressing the global challenge of student dropout, paving the way for more targeted and effective educational interventions. The findings also underscore the importance of reporting on the contribution of variables to model performance, as this information assists policymakers in identifying key areas of LMS log data that require attention. It is anticipated that the conclusions and suggestions made in this paper will work as a springboard for additional study and real-world implementations in the field of education to enhance student achievement and retention.

REFERENCES

Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*, *10*(1), 61–75. https://doi.org/10.1108/JARHE-09-2017-0113

Ahajjam, T., Moutaib, M., Aissa, H., Azrour, M., Farhaoui, Y., & Fattah, M. (2022). Predicting Students' Final Performance Using Artificial Neural Networks. *Big Data Mining and Analytics*, *5*(4), 294–301. https://doi.org/10.26599/BDMA.2021.9020030

Ahmed, E. (2024). Student Performance Prediction Using Machine Learning Algorithms. *Applied Computational Intelligence and Soft Computing*, *2024*. https://doi.org/10.1155/2024/4067721

Alsulami, A. A., AL-Ghamdi, A. S. A. M., & Ragab, M. (2023). Enhancement of E-Learning Student's Performance Based on Ensemble Techniques. *Electronics (Switzerland)*, *12*(6). https://doi.org/10.3390/electronics12061508

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. https://doi.org/10.3758/s13428-022-01939-9

Baker, R. S., Lindrum, D., Lindrum, M. J., & Perkowski, D. (n.d.). *Analyzing Early At-Risk Factors in Higher Education e-Learning Courses*.

Bañeres, D., Rodríguez-González, M. E., Guerrero-Roldán, A. E., & Cortadas, P. (2023). An early warning system to identify and intervene online dropout learners. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-022-00371-5

Bervell, B., & Umar, I. N. (2017). A decade of LMS acceptance and adoption research in Sub-Sahara African higher education: A systematic review of models, methodologies, milestones and main challenges. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(11), 7269–7286. https://doi.org/10.12973/ejmste/79444

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

Chung, J. Y., & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review*, *96*, 346–353. https://doi.org/10.1016/j.childyouth.2018.11.030

Dasi, H., & Kanakala, S. (n.d.). International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Student Dropout Prediction Using Machine Learning Techniques. In *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE* (Vol. 2022, Issue 4). www.ijisae.org

Figueroa-Canas, J., & Sancho-Vinuesa, T. (2020). Early prediction of dropout and final exam performance in an online statistics course. *Revista Iberoamericana de Tecnologias Del Aprendizaje*, *15*(2), 86–94. https://doi.org/10.1109/RITA.2020.2987727

Fu, Q., Gao, Z., Zhou, J., & Zheng, Y. (2021). CLSA: A novel deep learning model for MOOC dropout prediction. *Computers and Electrical Engineering*, *94*. https://doi.org/10.1016/j.compeleceng.2021.107315

Kamal, P., & Ahuja, S. (2019). An ensemble-based model for prediction of academic performance of students in undergrad professional course. *Journal of Engineering, Design and Technology*, *17*(4), 769–781. https://doi.org/10.1108/JEDT-11-2018-0204

Karim-Abdallah, B., Ayitey Junior, M., Appiahene, P., Harris, E., & Binful, D. K. (2025a). Application of Machine Learning Algorithms in Predicting Academic Performance of Students in Higher Education Institutes (HEIs): A Systematic Review and Bibliographic Analysis. *AFRICAN JOURNAL OF APPLIED RESEARCH*, *11*(1), 536–559. https://doi.org/10.26437/ajar.v11i1.869

Karim-Abdallah, B., & Harris, E. (2022). Modelling Customer Switching for Banks in Ghana. *Journal of Energy and Natural Resource Management (JENRM)*, *8*(1), 17–26. https://doi.org/10.26796/jenrm.v8i1.188

Karim-Abdallah, B., Okai Darko, G., Gyabaah, O., Oteng Fening, L., Chimsi, I., Derkyi, M. A. A., & Yeboah-Kyereh, A. (2025). Innovations, Technologies and Challenges Associated with Transnational Education. *AFRICAN JOURNAL OF APPLIED RESEARCH*, *11*(1), 560–587. https://doi.org/10.26437/ajar.v11i1.870

Mertova, Patricie., & Nair, C. Sid. (2011). *Student feedback : the cornerstone to an effective quality assurance system in higher education*. Chandos Publishing.

Palmer, S. (n.d.). *Modelling Engineering Student Academic Performance Using Academic Analytics**.

Porras, J. M., Lara, J. A., Romero, C., & Ventura, S. (2023). A Case-Study Comparison of Machine Learning Approaches for Predicting Student's Dropout from Multiple Online Educational Entities. *Algorithms*, *16*(12). https://doi.org/10.3390/a16120554

Prahani, B. K., Alfin, J., Fuad, A. Z., Saphira, H. V., Hariyono, E., & Suprapto, N. (2022). Learning Management System (LMS) Research During 1991–2021: How Technology Affects Education. *International Journal of Emerging Technologies in Learning*, *17*(17), 28–49. https://doi.org/10.3991/ijet.v17i17.30763

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

Saluja, R., Rai, M., & Saluja, R. (2023). Designing new student performance prediction model using ensemble machine learning. *Journal of Autonomous Intelligence*, *6*(1), 1–12. https://doi.org/10.32629/jai.v6i1.583

Tan, M., & Shao, P. (2015). Prediction of student dropout in E-learning program through the use of machine learning method. *International Journal of Emerging Technologies in Learning*, *10*(1), 11–17. https://doi.org/10.3991/ijet.v10i1.4189

Wang, Q., & Mousavi, A. (2023). Which log variables significantly predict academic achievement? A systematic review and meta-analysis. In *British Journal of Educational Technology* (Vol. 54, Issue 1, pp. 142–191). John Wiley and Sons Inc. https://doi.org/10.1111/bjet.13282

Yousafzai, B. K., Afzal, S., Rahman, T., Khan, I., Ullah, I., Rehman, A. U., Baz, M., Hamam, H., & Cheikhrouhou, O. (2021). Student-performulator: Student academic performance using hybrid deep neural network. *Sustainability (Switzerland)*, *13*(17). https://doi.org/10.3390/su13179775

Zhen, Y., Luo, J. Der, & Chen, H. (2023). Prediction of Academic Performance of Students in Online Live Classroom Interactions - An Analysis Using Natural Language Processing and Deep Learning Methods. *Journal of Social Computing*, *4*(1), 12–29. https://doi.org/10.23919/JSC.2023.0007

Zhou, Y., & Xu, Z. (2020). Multi-model stacking ensemble learning for dropout prediction in MOOCs. *Journal of Physics: Conference Series*, *1607*(1). https://doi.org/10.1088/1742-6596/1607/1/012004

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

Appendix A.

List of Primary Study Papers (PS) in this Literature

PS1. Abdelkader, H. E., Gad, A. G., Abohany, A. A., & Sorour, S. E. (2022). An Efficient Data Mining Technique for Assessing Satisfaction Level With Online Learning for Higher Education Students during the COVID-19. *IEEE Access*, *10*, 6286–6303. https://doi.org/10.1109/ACCESS.2022.3143035

PS2. Abdellaoui, B., Remaida, A., Sabri, Z., Abdellaoui, M., El Hafidy, A., El Bouzekri El Idrissi, Y., & Moumen, A. (2024). Analyzing emotions in online classes: Unveiling insights through topic modeling, statistical analysis, and random walk techniques. *International Journal of Cognitive Computing in Engineering*, *5*, 221–236. https://doi.org/10.1016/j.ijcce.2024.05.003

PS3. Abdullah, M., Al-Ayyoub, M., Shatnawi, F., Rawashdeh, S., & Abbott, R. (2023). Predicting students' academic performance using e-learning logs. *IAES International Journal of Artificial Intelligence*, *12*(2), 831–839. https://doi.org/10.11591/ijai.v12.i2.pp831-839

PS4. Abidi, S. M. R., Zhang, W., Haidery, S. A., Rizvi, S. S., Riaz, R., Ding, H., & Kwon, S. J. (2020). Educational sustainability through big data assimilation to quantify academic procrastination using ensemble classifiers. *Sustainability (Switzerland)*, *12*(15). https://doi.org/10.3390/su12156074

PS5. Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M., Bashir, M., & Khan, S. U. (2021). Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models. *IEEE Access*, *9*, 7519–7539. https://doi.org/10.1109/ACCESS.2021.3049446

PS6. 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*, *18*(11), 175–195. https://doi.org/10.3991/ijet.v18i11.37309

PS7. Ahmed, E. (2024). Student Performance Prediction Using Machine Learning Algorithms. *Applied Computational Intelligence and Soft Computing*, *2024*. https://doi.org/10.1155/2024/4067721

PS8. Alruwais, N. M. (2023). Deep FM-based predictive model for student dropout in online classes. IEEE Access, 11, 96954-96970.

PS9. Alruwais, N., & Zakariah, M. (2024). Student Recognition and Activity Monitoring in E-Classes Using Deep Learning in Higher Education. IEEE Access.

PS10. Alsabhan, W. (2023). Student Cheating Detection in Higher Education by Implementing Machine Learning and LSTM Techniques. *Sensors*, *23*(8). https://doi.org/10.3390/s23084149

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

PS11. Alsaidi, B. K., Ali, M. A., & Hussain, I. A. (2023). Study the Effect of Using Google Classroom on Academic Performance Under the Covid-19 Pandemic Using Data Mining Technique. *International Journal of Interactive Mobile Technologies*, *17*(6), 20–32. https://doi.org/10.3991/ijim.v17i06.38783

PS12. Arul, L., Rose, A. P. J., & Claral, M. T. (2022). An Early Intervention Technique for At-Risk Prediction of Higher Education Students in Cloud-based Virtual Learning Environment Using Classification Algorithms during COVID-19. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 13, Issue 1). www.ijacsa.thesai.org

PS13. Azizah, Z., Ohyama, T., Zhao, X., Ohkawa, Y., & Mitsuishi, T. (2024). Predicting at-risk students in the early stage of a blended learning course via machine learning using limited data. *Computers and Education: Artificial Intelligence*, *7*. https://doi.org/10.1016/j.caeai.2024.100261

PS14. Bañeres, D., Rodríguez-González, M. E., Guerrero-Roldán, A. E., & Cortadas, P. (2023). An early warning system to identify and intervene online dropout learners. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-022-00371-5

PS15. BENTALEB, A., & ABOUCHABAKA, J. (2022). ENSEMBLE LEARNING FOR MINING EDUCATIONAL DATA. Journal of Theoretical and Applied Information Technology, 100(9).

PS16. Chui, K. T., Fung, D. C. L., Lytras, M. D., & Lam, T. M. (2020). Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. *Computers in Human Behavior*, *107*. https://doi.org/10.1016/j.chb.2018.06.032

PS17. Colpo, M. P., Primo, T. T., & de Aguiar, M. S. (2024). Lessons learned from the student dropout patterns on COVID-19 pandemic: An analysis supported by machine learning. *British Journal of Educational Technology*, *55*(2), 560–585. https://doi.org/10.1111/bjet.13380

PS18. Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., & Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. *Decision Support Systems*, *135*. https://doi.org/10.1016/j.dss.2020.113325

PS19. Dasi, H., & Kanakala, S. (2022). Student dropout prediction using machine learning techniques. International Journal of Intelligent Systems and Applications in Engineering, 10(4), 408-414.

PS20. Dayanah Ayulani, I., Yunawan, A. M., Prihutaminingsih, T., Sarwinda, D., Ardaneswari, G., & Desjwiandra Handari, B. (2023). *Tree-Based Ensemble Methods and Their Applications for Predicting Students' Academic Performance*. *13*(3).

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

PS21. Ding, H., Chen, Y., & Wang, L. (2021). [Retracted] College English Online Teaching Model Based on Deep Learning. Security and Communication Networks, 2021(1), 8919320.

PS22. Figueroa-Canas, J., & Sancho-Vinuesa, T. (2020). Early prediction of dropout and final exam performance in an online statistics course. *Revista Iberoamericana de Tecnologias Del Aprendizaje*, *15*(2), 86–94. https://doi.org/10.1109/RITA.2020.2987727

PS23. Fu, Q., Gao, Z., Zhou, J., & Zheng, Y. (2021). CLSA: A novel deep learning model for MOOC dropout prediction. *Computers and Electrical Engineering*, *94*. https://doi.org/10.1016/j.compeleceng.2021.107315

PS24. Gligorea, I., Yaseen, M. U., Cioca, M., Gorski, H., & Oancea, R. (2022). An Interpretable Framework for an Efficient Analysis of Students' Academic Performance. *Sustainability (Switzerland)*, *14*(14). https://doi.org/10.3390/su14148885

PS25. Guy, X. R., Byrne, B., & Dobos, M. (2018). Optional anatomy and physiology e-learning resources: student access, learning approaches, and academic outcomes. *Adv Physiol Educ*, *42*, 43–49. https://doi.org/10.1152/advan.00007.2017.-Anatomy

PS26. Hamim, T., Benabbou, F., & Sael, N. (2021). Survey of Machine Learning Techniques for Student Profile Modelling. *International Journal of Emerging Technologies in Learning*, *16*(4), 136–151. https://doi.org/10.3991/ijet.v16i04.18643

PS27. He, Y., Chen, R., Li, X., Hao, C., Liu, S., Zhang, G., & Jiang, B. (2020). Online at-risk student identification using RNN-GRU joint neural networks. *Information (Switzerland)*, *11*(10), 1–11. https://doi.org/10.3390/info11100474

PS28. Holicza, B., & Kiss, A. (2023). Predicting and Comparing Students' Online and Offline Academic Performance Using Machine Learning Algorithms. *Behavioral Sciences*, *13*(4). https://doi.org/10.3390/bs13040289

PS29. Huang, Q., & Chen, J. (2024). Enhancing academic performance prediction with temporal graph networks for massive open online courses. *Journal of Big Data*, *11*(1). https://doi.org/10.1186/s40537-024-00918-5

PS30. Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2020). *Multi-split Optimized Bagging Ensemble Model Selection for Multi-class Educational Data Mining*. https://doi.org/10.1007/s10489-020-01776-3

PS31. Jawthari, M., & Stoffová, V. (2021). Predicting students' academic performance using a modified kNN algorithm. *Pollack Periodica*, *16*(3), 20–26. https://doi.org/10.1556/606.2021.00374

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

PS32. Kabathova, J., & Drlik, M. (2021). Towards predicting student dropout in university courses using different machine learning techniques. *Applied Sciences (Switzerland)*, *11*(7). https://doi.org/10.3390/app11073130

PS33. Kostopoulos, G., Kotsiantis, S., Fazakis, N., Koutsonikos, G., & Pierrakeas, C. (2019). A Semi-Supervised Regression Algorithm for Grade Prediction of Students in Distance Learning Courses. *International Journal on Artificial Intelligence Tools*, *28*(4). https://doi.org/10.1142/S0218213019400013

PS34. Kukkar, A., Mohana, R., Sharma, A., & Nayyar, A. (2023). Prediction of student academic performance based on their emotional well-being and interaction on various e-learning platforms. Education and Information Technologies, 28(8), 9655-9684.

PS35. Liu, Y., Huang, Z., & Wang, G. (2023). Student learning performance prediction based on online behavior: an empirical study during the COVID-19 pandemic. *PeerJ Computer Science*, *9*. https://doi.org/10.7717/peerj-cs.1699

PS36. Manikandan, S., & Chinnadurai, M. (n.d.). *Proofs Evaluation of Students' Performance in Educational Sciences and Prediction of Future Development using TensorFlow**.

PS37. Masood, J. A. I. S., Kalyan Chakravarthy, N. S., Asirvatham, D., Marjani, M., Abdulkareem Shafiq, D., & Nidamanuri, S. (2024). A Hybrid Deep Learning Model to Predict High-Risk Students in Virtual Learning Environments. *IEEE Access*, *12*, 103687–103703. https://doi.org/10.1109/ACCESS.2024.3434644

PS38. Memon, M. Q., Lu, Y., Yu, S., Memon, A., & Memon, A. R. (2022). The Critical Feature Selection Approach Using Ensemble Meta-Based Models to Predict Academic Performances. *International Arab Journal of Information Technology*, *19*(Special Issue 3A), 523–529. https://doi.org/10.34028/iajit/19/3A/12

PS39. Mi, H., Gao, Z., Zhang, Q., & Zheng, Y. (2022). Research on Constructing Online Learning Performance Prediction Model Combining Feature Selection and Neural Network. *International Journal of Emerging Technologies in Learning*, *17*(7), 94–111. https://doi.org/10.3991/ijet.v17i07.25587

PS40. Moreno-Marcos, P. M., Muñoz-Merino, P. J., Maldonado-Mahauad, J., Pérez-Sanagustín, M., Alario-Hoyos, C., & Delgado Kloos, C. (2020). Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs. *Computers and Education*, *145*. https://doi.org/10.1016/j.compedu.2019.103728

PS41. Nayak, P., Vaheed, S., Gupta, S., & Mohan, N. (2023). Predicting students' academic performance by mining the educational data through machine learning-based classification model. Education and Information Technologies, 28(11), 14611-14637.

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

PS42. Nazempour, R., & Darabi, H. (2023). Personalized Learning in Virtual Learning Environments Using Students' Behavior Analysis. *Education Sciences*, *13*(5). https://doi.org/10.3390/educsci13050457

PS43. Olaniyan, D., Olaniyan, J., Obagbuwa, I. C., Esiefarienrhe, B. M., & Bernard, O. P. (2024). Parallel Attention-Driven Model for Student Performance Evaluation. *Computers*, *13*(9). https://doi.org/10.3390/computers13090242

PS44. Pecuchova, J., & Drlik, M. (2023). Predicting Students at Risk of Early Dropping Out from Course Using Ensemble Classification Methods. *Procedia Computer Science*, *225*, 3223–3232. https://doi.org/10.1016/j.procs.2023.10.316

PS45. Porras, J. M., Lara, J. A., Romero, C., & Ventura, S. (2023). A Case-Study Comparison of Machine Learning Approaches for Predicting Student's Dropout from Multiple Online Educational Entities. *Algorithms*, *16*(12). https://doi.org/10.3390/a16120554

PS46. Poudyal, S., Mohammadi-Aragh, M. J., & Ball, J. E. (2022). Prediction of Student Academic Performance Using a Hybrid 2D CNN Model. *Electronics (Switzerland)*, *11*(7). https://doi.org/10.3390/electronics11071005

PS47. Prenkaj, B., Distante, D., Faralli, S., & Velardi, P. (2021). Hidden space deep sequential risk prediction on student trajectories. *Future Generation Computer Systems*, *125*, 532–543. https://doi.org/10.1016/j.future.2021.07.002

PS48. Qiu, F., Zhang, G., Sheng, X., Jiang, L., Zhu, L., Xiang, Q., Jiang, B., & Chen, P. Kuo. (2022). Predicting students' performance in e-learning using learning process and behaviour data. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-021-03867-8

PS49. Rawat, S., Kumar, D., Kumar, P., & Khattri, C. (2021). A systematic analysis using classification machine learning algorithms to understand why learners drop out of MOOCs. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-021-06122-3

PS50. Saleem, F., Ullah, Z., Fakieh, B., & Kateb, F. (2021). Intelligent decision support system for predicting student's e-learning performance using ensemble machine learning. *Mathematics*, *9*(17). https://doi.org/10.3390/math9172078

PS51. Santamaría-Lopez, T., Patiño-Perez, D., González-Ruiz, V., & Flores-Carvajal, L. (2023). Implementation of machine learning techniques and creation of an artificial neural network for the prediction of the academic performance of students in university environments that use e-learning and streaming. Movilidad para la descarbonización, 282.

PS52. Sathe, M. T., & Adamuthe, A. C. (2021). Comparative study of supervised algorithms for prediction of students' performance. *International Journal of Modern Education and Computer Science*, *13*(1), 1–21. https://doi.org/10.5815/ijmecs.2021.01.01

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

PS53. Tan, M., & Shao, P. (2015). Prediction of student dropout in E-learning program through the use of machine learning method. *International Journal of Emerging Technologies in Learning*, *10*(1), 11–17. https://doi.org/10.3991/ijet.v10i1.4189

PS54. Vaarma, M., & Li, H. (2024). Predicting student dropouts with machine learning: An empirical study in Finnish higher education. *Technology in Society*, *76*. https://doi.org/10.1016/j.techsoc.2024.102474

PS55. Velasco, C. L. R., Villena, E. G., Ballester, J. B., Prados, F. Á. D., Alvarado, E. S., & Álvarez, J. C. (2023). Forecasting of Post-Graduate Students' Late Dropout Based on the Optimal Probability Threshold Adjustment Technique for Imbalanced Data. *International Journal of Emerging Technologies in Learning*, *18*(4), 120–155. https://doi.org/10.3991/ijet.v18i04.34825

PS56. Villegas-Ch, W., García-Ortiz, J., & Sánchez-Viteri, S. (n.d.). *Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. Application of Artificial Intelligence in Online Education: Influence of Student Participation on Academic Retention in Virtual Courses*. https://doi.org/10.1109/ACCESS.2017.DOI

PS57. 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

PS58. Yousafzai, B. K., Afzal, S., Rahman, T., Khan, I., Ullah, I., Rehman, A. U., Baz, M., Hamam, H., & Cheikhrouhou, O. (2021). Student-performulator: Student academic performance using a hybrid deep neural network. *Sustainability (Switzerland)*, *13*(17). https://doi.org/10.3390/su13179775

PS59. Zhang, J., Qiu, F., Wu, W., Wang, J., Li, R., Guan, M., & Huang, J. (2023). E-Learning Behavior Categories and Influencing Factors of STEM Courses: A Case Study of the Open University Learning Analysis Dataset (OULAD). *Sustainability (Switzerland)*, *15*(10). https://doi.org/10.3390/su15108235

PS60. Zhang, X. (2024). A Decision Tree and Generalised Regression Neural Network Based Assessment Model for Online Education.

PS61. Zhen, Y., Luo, J. Der, & Chen, H. (2023). Prediction of Academic Performance of Students in Online Live Classroom Interactions - An Analysis Using Natural Language Processing and Deep Learning Methods. *Journal of Social Computing*, *4*(1), 12–29. https://doi.org/10.23919/JSC.2023.0007

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

Appendix B

Table 4: Data Extraction Sheet

PS	Data Size (No. of Students)	No. of Attributes or features used in modelling	Significant attribute or feature in the proposed model	Algorithms	Chosen Algorithm	Results			
						Accuracy	Precision	Recall	F1
PS1	18691	20	Attractiveness of the lectures	KNN, SVM	KNN	1			
PS2	1000	Images	Facial Emotion Recognition (FER)	CNN, Non-Negative Matrix Factorization (NMF)	CNN				
PS3	4874	9	Quiz event	Random Forest, Bayesian Ridge, Adaboost, XGBoost and Voting Model.	Voting Algorithm	Best RMSE: 2.08 on HSS129.			
PS4	4094	7	Total Hints Given	Logistic regression, Decision Tree, Gradient Boosting, Random Forest	Gradient Boosting	0.9536	0.9173	0.978	0.947
PS5	32593	31	Student Assessment scores	RF, SVM, KNN, Extra Tree, Adaboost, Gradient Boosting, Deep Feed Forward Neural Network	RF	0.9100	0.9200	0.910	0.910

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

PS6	6450	24	partial grades from the first and second grading periods	XGBoost, RF, SVM, K-Means clustering	RF	0.967	0.9751	0.9892	0.98
PS7	32005	7		KNN, SVM, DT, NB	SVM	0.9603	0.9418	0.9843	
PS8	641138	3 PCA		DeepFM	DeepFM	0.9910	0.9820	0.9950	0.9870
PS9	20			CNN	CNN	0.9910	0.9830	0.9720	0.9910
PS10	94	6		LSTM	LSTM	0.9200	0.6200	0.5200	0.7800
PS11	219	44	Lack of knowledge of technical matters	CNN	CNN	0.9800	0.9700	0.9520	0.9810
PS12	530	46	CGPA	kNN, SVM, LDA, RF,	RF	0.8861	0.5667	0.9607	
PS13	424	12	Study in time	RF	RF	0.8240	0.3530	0.1760	0.2350
PS14	957			DT	DT	0.9071	0.9824	0.6378	0.7025
PS15	480	33		LR, NB, SVM, RF, XGBoost, Voting	Voting Classifier	Data1: 0.85 / Data2: 0.92			
PS16	32593	31		RTV-SVM	RTV-SVM	0.938	0.936	0.9400	
PS17	3371	67		DT	DT	0.8539	0.8344	0.8248	
PS18	10554	122	Academic Engagement	LR, SVM, NN, DT, RF, Bagging, Boost, Logistic Model Tree, Logit leaf model (LLM)	Logit Leaf Model	AUC = 0.839	TLD =1.798		
PS19	261	10	No. of times a student accesses a course.	LR, NB, DT, SVM, RF	LR	0.930	0.960	0.9800	0.9600

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

PS20	232	8		RF, XGBoost, LGBM	LGBM	0.9260	0.483	0.868	
PS21	64	4		CNN, MLP, Kmeans	CNN	0.9830			
PS22	197	12		DT	DT		0.751	0.703	0.7200
PS23	79186	7		SVM, LSTM, Grated Recurrent Unit, CNN, CnnLstmStaticAction model	CLSA	0.8760	0.874	0.865	0.8690
PS24	32593	7	Total number of clicks	LR, Lasso regression, RF, DT, SVR, ANN, Gradient Boosting Regressor	GBR	0.9500	0.940	0.830	0.8700
PS25	137	10	Video Clips	Statistical learning					
PS26	480	16		NB, ANN, SVM, DT, RF, Kmeans. KNN, LR	ANN	0.7778			
PS27	32593	20	Assessment performance history	RNN, GRU, LSTM	RNN	0.9000		0.850	
PS28	145	33		SVM, DT, RF, KNN	RF	0.9900	0.990	1	0.9900
PS29	3897	7		AppTGN, APPGRU	APPTGN	0.8320		0.9021	0.7311
PS30	486	5	Assignment 01	KNN, RF, SVM, LR, NB, NN	RF	0.667, 0.931	0.53, 0.71	0.43, 0.73	-, 0.720
PS31	480	16		Standard KNN, Voting	Jaccard distance vote	0.8020			
PS32	261	3	No. of times a student accesses a	LR, DT, RF, NB, SVM, NN	RF	0.9300	0.81	0.96	0.910

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

			course during a period						
PS33	1073	91		MSSRA, LR, SMOreg, kNN, M5 Rules, M5 Model Tree, RF	MSSRA	Regression: RMSE = 7.625			
PS34				ANN, LSTM, RNN, CNN, SVM, DT, NB, RF	Combined (LSTM, RF, B)	0.96			
PS35	508	6		RF, NB, LibSVM, MLP, SMO, J48, J48graft, Adaboost, Bagging, Optimized forest, MOEFC, Voting	Adaboost	0.92105	0.9240	0.9210	0.9220
PS36	2500	9		DT, Kmeans, k-mediots, NB, SVM, CNN	CNN	0.85 - 0.90			
PS37	32593	20		HDL, MLP, DFFNN	HDL	0.9567	0.9424	0.9645	0.9378
PS38	11814	18	family	DT, RF, DL, LR, k-star, Ensemble Boosting, Bagging, Adaboost)	K-Star	0.92			0.8100
PS39	48000	6	Interaction times	LR, DT, NB, SVM, DNN	DNN with MLP	0.97	0.9880	0.9970	0.9930
PS40	25706	34	Percentage of completed assignments	RF, GLM, SVM, DT	RF	AUC = 0.96			
PS41		6	Students Behaviour	Opt-MLP, DT, NB,RF, MLP	Opt-MLP	0.9074			

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

PS42	32593	17		LR, RF, NB, k-NN, LDA, QDA, DT, SVM, Ridge classifier, Gradient boosting	Gradient boosting	0.7944	0.774	0.871	0.8197
PS43	40,000,000	7		LSTM, Multitask learning	LSTM	1	1	1	1
PS44	110	22	test sA1-A3	LR, RF, DT, kNN, SVM, Adaboost, XGBoost	Adaboost	0.9706	0.97	0.95	0.9600
PS45	337,988	11		Transfer learning, voting, stacking,	Repurposed transfer		AUC = 0.8837		
PS46	32593	37		2D-CNN	2D-CNN	0.88			
PS47	5000	97		GRU-AE, kNN, DNN, LR, DT, RF, SVM, MC	GRU-AE	AUC= 0.79			0.7520
PS48	6272	12		SVC, NB, kNN, Softmax	NB	0.9165			0.9423
PS49	367375	12	Course content	NN, DT, LR, Gradient boosting, DL	NN	0.9697			
PS50	480	17	Visited resources	DT, RF, Gradient Boosting, NB, kNN, Ensemble (Bagging, Boosting, Stacking, Voting)	Ensemble Stacking	0.8195	0.8199	0.8195	0.8195
PS51	1248	6		RF, SVM, KNN, LR, ANN, XGBoost	ANN	0.824	0.824	0.9730	0.8930
PS52	395	12	G1and G2 scores	MLP (J48, C5.0), CART, NB, kNN, RF, SVM	MLP (C5.0)	0.9333	0.935	0.9190	
PS53	62375	26		ANN, DT, Bayesian Networks	DT	0.9463	0.6389	0.9576	0.7191

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 2 (2025), pp.278-312
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i2.1038
Received: November 27, 2024
Peer reviewed: January 15, 2025
Revised: February 22, 2025
Published: April 2025

PS54	8813	14	Accumulated Credits	CatBoost, SVM, XGBoost, RF, LR, NB, NN, kNN, LDA	CatBoost	0.721	0.810	0.470	0.5900
PS55	10934	12	Duration Subjects	LR, RF, NB, NN	NN	0.8	0.590	0.610	0.6000
PS56	301	3	Student participation	ARIMA, kMeans, Prophet, DBSCAN	Prophet		0.920	0.880	0.9000
PS57	32593	30	Active participation	SVM, LR, Deep Ann(MLP)	MLP	0.8448	0.860	0.617	
PS58	1044	33	Grade of the second period	SVM, LR, kNN, Multinomial NB, RF, CNN, LSTM, Bi-LSTM	Bi-LSTM	0.9016	0.900	0.900	0.9000
PS59	21403	19	Learning preparation behaviours	kNN dis, NB, SVM(SVC rbf), DT,	SVM(SVC rbf)	0.9044	0.903		0.9352
PS60	30	15		C4.5 DT, Generalized Regression Neural Network (GRNN)	GRNN	Relative error of predictio n			
PS61	89694	24		DT, ANN, CNN	CNN	0.84	0.86	0.810	0.8300

ISSN: 2408-7920