African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

APPLICATION OF MACHINE LEARNING ALGORITHMS IN PREDICTING ACADEMIC PERFORMANCE OF STUDENTS IN HIGHER EDUCATION INSTITUTES (HEIS): A SYSTEMATIC REVIEW AND BIBLIOGRAPHIC ANALYSIS

Karim-Abdallah, B.[1], Ayitey Junior, M.[2], Appiahene, P.[3], Harris, E.[4] and Binful, D. K.[5]
[1]*Quality Assurance and Academic Planning Directorate, University of Energy and Natural Resources, Sunyani, Ghana.*
[2&3]*Department of Information Technology and Decision Science, University of Energy and Natural Resources, Sunyani, Ghana.*
[4]*Department of Statistics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.*
[5] *Directorate of Information Technology, University of Energy and Natural Resources, Sunyani, Ghana.*
[1]*bright.karim-abdallah@uenr.edu.gh*

ABSTRACT
Purpose: This study aims to identify trends, key research areas, popular predictive features, machine learning algorithms used, and gaps in the existing literature.
Design/Methodology/Approach: This study selected 60 articles published between 2018 and August 31, 2023, from Google Scholar and Scopus databases to address the identified knowledge gap. A systematic literature review and bibliographic analysis were conducted using both quantitative and qualitative approaches, with PRISMA chosen as the reporting format.
Findings: The study reveals that machine learning models, particularly Decision Trees, followed by Random Forests, Artificial Neural Networks, Support Vector Machines, and Naïve Bayes, have significantly contributed to predicting student academic performance. The datasets predominantly utilised by researchers include students' academic records, demographics, activities, and behaviour, which are crucial for predicting academic performance and evaluating model effectiveness.
Research Limitation: These limitations highlight the challenges in conducting comprehensive reviews of machine learning applications in educational contexts while acknowledging the evolving nature of both technology and educational assessment methods.
Practical Implication: This research has the potential to inform evidence-based decision-making, promote personalised learning experiences and enable early interventions for at-risk students
Social Implication: The study underscores the potential of machine learning algorithms to promote equity and inclusion, provide targeted student support, empower individuals, raise ethical considerations, foster community engagement, and support lifelong learning initiatives.
Originality and Value: This research uniquely explores machine learning applications to predict student academic performance, identifying classroom participation and examination scores as key predictors. It offers a pragmatic approach to educational practice, identifies future research opportunities, and advocates for data-driven decision-making in higher education.
***Keywords**: Academic performance. bibliographic analysis. machine learning. prediction. students*

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

INTRODUCTION

Higher Education Institutions (HEIs) are increasingly under pressure to improve their student's academic performances, which will significantly impact their graduates' employability as a measure of their study programme's effectiveness. HEIs are expected to give future learning opportunities and to assist the globe in addressing the fast-evolving social, cultural, economic, and environmental sustainability concerns of the twenty-first century (Mertova & Nair, 2011).This expectation ensures the achievement of Sustainable Development Goal 4 (SDGs), which states, "ensure inclusive, equitable quality education, and promote lifelong learning opportunities for all." The above can be achieved when the university's products are well-prepared to meet the challenging demands of an unknown future.

To make students benefit fully from HEIs, their academic performances must be the priority of the Institute. Predictive modelling is one of the prominent techniques that can monitor the academic performance of students (Addo & Sackey, 2022; Aliyu Sani & Haruna, n.d.; Karim-Abdallah & Harris, 2022; Sackey & Ngewana, 2022).Predicting students' academic performances makes it possible to identify significant features or variables in determining their academic performance. Academics have used different machine learning (ML) models with other datasets to predict students' academic performance (SAP) in HEIs.

Duong et al., (2023) used a Support Vector Machine (SVM) technique with the Vietnamese Universities dataset to predict academic performances and Mohd Talib et al., (2023) used K-Means Clustering and Support Vector Machine with students' behavioural patterns to classify the academic performances of their respondents. Alsulami et al., (2023) used Decision Trees (DT), Naïve Bays (NB), and Random Forest (RF) with student's E-learning data to analyse and predict student performance and Dayanah Ayulani et al. (2023) used RF, EG Boosting and LG boosting Machine to identify students who are at likely to fail their courses.

Most researchers have used ML techniques to predict and classify the student's academic performance with a variety of datasets and report on the evaluation metrics of the model; however, they failed to identify the most significant features or variables of their datasets contributing to their model performance and come out with pragmatic measures to help the underperforming or students at risk. This study's objective is to provide a complete assessment of the body of literature on the use of ML for modelling students' academic performance and identifying the significant features or variables used in the datasets in HEIs. Factors impacting students' academic performances can provide insights into which teaching methods are most effective.

A systematic literature review (SLR) of applications of ML techniques in classifying or predicting students' academic performance in HEIs aims to explore the various ways in which ML techniques have been utilised to predict or classify student performance.

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

THEORIES UNDERPINNING THE STUDY

This section presents the literature on applying ML techniques to predict students' academic performances in HEIs. It also includes work that shows how ML approaches were used to classify or predict students' academic performances and literature that exclusively shows significant features or variables in the dataset in these approaches or techniques. In addition, relevant SLRs on our subject are discussed.

Machine learning for prediction of academic performances of higher education institutions

ML is becoming increasingly popular because computers are faster, and the amount of data generated and stored is growing exponentially. Machine learning techniques and algorithms have evolved into strong instruments for studying vast and complicated data sets, supporting scientists in significant advances in various domains of science and industry. (Russell, 2018). The activity of teaching computers to learn from data is known as machine learning. There are different types of machine-learning systems: Supervised, Unsupervised, Semi-supervised, and Reinforcement Learning (Cielen et al., 2016.; Russell, 2018). Supervised machine learning techniques aim to detect findings and learn by looking for patterns in a labelled data set.

Human interaction is required to label the data. Examples of supervised algorithms are K-Nears Neighbours, Linear Regression, Neural Networks, SVM, Logistic Regression, DTs, and RF. Unsupervised machine learning approaches do not rely on labelled data and aim to detect patterns in data collection without human intervention. Examples of Unsupervised machine learning algorithms are Clustering (k-means, hierarchical cluster analysis), Association rule learning (Eclat, apriori), and Visualization and dimensionality reduction (kernel PCA, t-distributed, PCA). Semi-supervised learning approaches require labelled data and, thus, human engagement to detect patterns in the data set. However, they can still proceed towards a result and learn even when given unlabeled data. Another form of machine-learning system is reinforcement learning. An agent's "AI system" will observe its surroundings, take prescribed actions, and be rewarded. The agent must learn on its own with this type. (Cielen et al., 2016).

Many researchers are now making use of ML and AI in different domains to make informed decisions (Afolagboye et al., 2023; Ahmed et al., 2021; Ojokoh et al., 2020). The tertiary education sector is one area where data can be effectively used to make data-driven decisions. Identifying factors or features in the dataset that contribute significantly to modelling students' academic performance will help HEIs and other stakeholders know where to channel limited resources to save the situation.

Xue and Niu (2023) used multi-output-based hybrid integrated models with Superstar Learning Communication Platform (SLCP) data for students' performance prediction. They used different ensemble techniques to model the SLCP data. Their results show that using the first six weeks of SLCP data to predict students' performance, XGBoost and gradient-boosted trees (GDBT) outperformed the

ISSN: 2408-7920

other ensemble methods. The significant features of the dataset which contributed to the model's performance were not named.

Alruwais and Zakariah (2023) employed seven different classifiers: SVM, LR, RF, DT, Gradient Boosting Machine (GBM), Gaussian Naive Bayes (GNB), and Multi-Layer Perceptron (MLP) to predict students' academic performances using a dataset obtained from an e-learning environment which consists of six attributes or features: the goal object's degree of study time, degree of study counts for the goal object, degree of study time for the related object, learning percentage for the related objects, performance in exams for goal object and user knowledge state. These seven attributes were categorised into three major categories: individual behavioural, exam score-related, and knowledge level. In their findings, the GBM exhibited the highest prediction accuracy and outperformed the other classifiers. The contribution of various attributes to the model performances was not identified.

Alsulami et al. (2023) used e-learning data with DT, NB, and RF, further enhanced using three ensemble techniques: bagging, boosting, and voting to predict students' performance. The dataset consists of demographic attributes such as place of birth, nationality, gender, parent responsibility for their children, academic attributes such as educational stage and grade level, and behavioural attributes including opening resources and raising hands in class. Their objective was to identify the factors in the dataset which affect students' academic performance. They concluded that the DT method with boosting performed better than the other methods. The contribution of each attribute to the model was not declared. Abou Naaj et al. (2023) used a fuzzy-neural approach to identify the factors impacting student academic achievement and create a model that predicts and explains variations in course grades among students using course category, student course attendance rate, gender, high-school grade, school type, grade point average (GPA), and course delivery mode as inputs features. In their study, the authors concluded that the most significant variables or features of students' academic achievement are GPA, course category and course delivery.

Despite machine learning's advantages and the strides it has made in recent years, it seems difficult for academics to determine the most effective technique and how to identify variables or factors that have the most significance on modelling performance when predicting student academic performance. Several studies have already used machine learning to predict student academic performance in HEIs. However, there is a need for collected literature that will address the domain's issues, challenges, and future research aims.

Related works

All academic institutions now prioritise student success, necessitating the study of the factors that influence that achievement. Machine learning algorithms have generated considerable attention in the scientific community due to their potential to predict or classify student performance.

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

Issah et al. (2023b) undertook a study on the application of ML to identify the factors affecting students' academic performance. They uncovered various machine-learning approaches after thoroughly studying the literature with 84 selected papers. The study demonstrates how researchers have accomplished pattern-mapping student traits and their impact on academic achievement. How the total research coverage of student attributes and the ML approaches are used to predict students' performance was attempted to be answered.

According to the research, the analysis of the 84 publications shows that the following features: academic performance (grade point average, grade level, high school score, attendance to lessons, attendance to lessons, number of courses per semester), demographic (gender, nationality, place of birth, age), behavioural (Raised hands, visit resources, school satisfaction, discussion, attend class, answer questions), psychological (personality, motivation, learning strategies, approach to learning, contextual influence), family background (Mother & father Education, family income, location of parents), and school environment (school size, medium of instruction, lecturer/teacher behaviour in class) were used in modelling students' academic performance.

Among these features, academic and demographic variables are the dominant characteristics mainly employed in student academic performance prediction. The survey demonstrates that classification and decision trees are the most popular techniques and algorithms. The review also identifies demographic and practical knowledge gaps due to the paucity of research on fundamental academic performance and the lack of intervention approaches for preventing low performance by connecting these important factors to student achievement. The extent of the population sample requires a benchmarked dataset and the integration of suitable intervention outlines that will map the learner's performance early in their academic career to close these perceived gaps. In all papers under review, the study never mentioned significant variables that contributed to model performance.

In a study by Roslan and Chen (2022), they examined 58 articles published from 2015 to 2021. They aimed to uncover elements and methodologies employed in predicting students' academic performance and trends in academic publications. The results reveal that current research primarily focuses on identifying factors that influence student performance and the effectiveness of machine learning algorithms. Additionally, it highlights that student academic records and demographics play a role in determining their performance. Among the machine learning approaches, classification is most frequently utilised, with the Decision Tree classifier being the preferred algorithm. While they successfully identified factors used by researchers to predict students' academic performance, they could not assess the impact of these factors on model performance.

According to Sekeroglu et al. (2021), some of the main goals of education include enhancing quality, developing and implementing strategies that help students, and anticipating students' achievement throughout the term, after the term, or in the future. AI and ML are technologies utilised in this discipline

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

to attain the intended aims due to their distinct capacity to forge associations and produce correct outcomes. However, the variety of studies and the disparities in their subject matter led to confusion and lessened their capacity to lead new research. They conducted a thorough literature assessment of research that predicted student performance between 2010 and 2020 in three separate databases as part of the study. They were then shown as percentages by classifying their findings as either model, dataset, validation, evaluation, or goals. Critical gaps, potential fixes, and similarities and variations between the research are provided. The findings and gaps might be filled with systematic assessment and validation techniques. It is decided that future research on student performance classification or prediction should concentrate more on deep learning models. Finally, the advantages and challenges of a worldwide dataset produced by a global consortium of education information are discussed.

METHODOLOGY

A systematic literature review uses a research approach that must be neutral and complete to evaluate all existing research on the given topic (Kitchenham, 2021). The protocol's structure includes planning, conducting, and reporting.

The first two sections deal with planning, whereas section three deals with reporting.

Planning

Planning is the first stage in SLR. It consists of five stages described in more depth in the subsections below.

Determine research questions: This study utilised the standards offered by (Kitchenham, 2021) in formulating its research questions.

RQ1: What are the publication trends of machine learning methods in predicting and improving academic performances in HEIs,

RQ2: What features dominate the dataset used in predicting the academic performances of students in HEIs?

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

RQ4: What ML methods are used in classifying or predicting the academic performances of students in HEIs?

RQ5: What gaps exist in the current literature?

Decide on keywords.

The research questions were used to choose the search terms. The search terms in this study are**:**

("machine learning") AND ("prediction" OR "classification" OR "influencing") AND ("improving" OR "enhancing") AND ("students academic performance" OR "success of students") AND ("higher education" OR "Tertiary") AND PUBYEAR > 2020 AND PUBYEAR < 2024 AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (EXACTKEYWORD, "Educational Data Mining")OR LIMIT-TO(EXACTKEYWORD , "Forecasting") OR LIMIT-TO (EXACTKEYWORD , "Data Mining") OR LIMIT-TO (EXACTKEYWORD , "Machine Learning") OR LIMIT-TO (EXACTKEYWORD, "Students"))

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

Identify the source.

The Scopus and Google Scholar databases (https://scholar.google.com/) offer thorough, well-vetted information on various subjects.

Determine the criteria for inclusion and exclusion.

One of the quality assurance tools employed is inclusion and exclusion criteria. The inclusion criteria include journal articles, primary literature, publications between 2018 and August 31, 2023, meeting the research keywords, classified as ML studies, availability and accessibility in the database, and papers written in English. The exclusion criteria include conference papers, books, magazines, editorials, secondary literature, tertiary literature and duplicate documents in the database. After being pulled from the search results, all articles that matched the specified criteria were included in this SLR.

Select a data extraction method.

The next step is to gather data from each research by doing database searches using the predefined search term. Journal articles that did not have the correct information for one or more areas were eliminated from this study.

Conducting the review

The second part of this SLR technique focuses on conducting the review. This phase is divided into four segments, which are as follows.

Identify the research.

In this phase, the search string determined in “decide on the key string” was deployed to search the Scopus and Google Scholar online databases. For 392 primary papers, 242 papers were retrieved from the Scopus database and 150 from Google Scholar.

Selecting the research.

The literature review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The PRISMA flowchart is shown in Figure 1. This SLR analysed only relevant journal papers that met the inclusion criteria and examined the substance of each selected article to ensure its inclusion. The publications were evaluated, and 120 were rejected because they did not fit the inclusion requirements or were unrelated to the study's subject. The screening, which included semi-automatic document selection by examining the entire text of the remaining publications, also revealed that 117 papers were not ML research. Furthermore, 58 publications did not provide the whole text, and 37 papers were unrelated to the study's subject. These articles were also eliminated from this investigation. As a result, this SLR only looked at 60 journal papers. The list of all the chosen articles is in the appendix.

ISSN: 2408-7920

Arca Academic Publisher

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

Synthesising the data.

This SLR retrieved three research emphasis categories from the 60 selected articles. These include identifying factors impacting student performance, the performance of ML algorithms, and ML about e-learning systems, as shown in Table 1. The study then identified six separate kinds of characteristics influencing student performance. Table 2 lists and describes the categories of factors that affect student performance. This study identified three ML techniques based on the 89 ML algorithms employed to predict student performance. Table 3 shows the ML method categories for the selected articles. Figure 1 shows 242 articles on the Scopus database and 150 on Google Scholar with the above query string. After searching databases and identifying possibly relevant articles, the first step is removing duplicate articles (because the same article may be located in multiple databases).

Table 1: Research focus of the selected articles

Research Topic	Paper ID
Identification of attributes influencing student's performance	P13,P18,P20,P24,P25,P26,P29,P30,P33,P38,P41,P44,P45,P46,P47,P56
Machine learning performance	P1,P2,P3,P4,P6,P8,P9,P10,P11,P12,P15,P16,P17,P19,P22,P23,P28,P31,P34,P35, P36,P37,P39,P40,P42,P43,P45,P49,P50,P51,P52,P53,P54,P55,P57,P58,P59,P60
E-learning	P5,P7,P13,P14,P21,P27,P32,P48

ISSN: 2408-7920

Arca Academic Publisher

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

Table 2: Categories of factors that affect the student's performance and their description

Attributes affecting the student's academic success	Description	Paper ID
Students' Academic Record	The records related to academic performance e.g., study duration, student's exam result, assessment mark, GPA, CGPA, and class attendance.	P2,P3,P4,P7,P9,P10,P12,P13,P16,P19, P21,P22,P23,P24,P27,P28,P29,P30,P31, PP33,P35,P36,P37,P38,P39,P41,P42,P43, P46,P47,P51,P52,P54,P56,P57,P58,P59
Students' demographics	The demographic factors include age, gender, geographical affiliation, ethnicity, nationality, marital status, socioeconomic status,(SES), parental education, language, financial, and religious affiliations.	P1,P6,P7,P8,P12,P13,P15,P16,P17,P18, P20,P23,P24,P25,P30,P31,P32,P33,P35, P36,P39,P40,P41,P42,P43,P44,P46,P49, P53,P54,P55,P56,P57,P58,P60
Students' activities	Information that related to the student's activities such as extracurricular activities, e-Learning, activities, internet usage, and timespent with friends.	P4,P5,P7,P8,P12,P15,P16,P17,P21, P25,P34,P40,P43,P46,P48,P54
Student's behaviors	The way that the students behave e.g., procrastination and self-discipline.	P2,P3,P12,P13,P14,P15,P18,P20, P24,P27,P25,P28,P31,P33,P35,P46
Students' motivation	Factors that make the students be motivated, e.g., parents' involvement and students' interest.	Null
Socio-economic background	Economic background of students and parents	P26,P30,P33,P34,P35,P39,P41,P42, P44,P53,P55

Table 3: ML approaches and articles

Machine learning Approaches	Article ID
Classification	P1, P2, P3, P4,P5,P6,P7,P8,P9,P10,P11, P12,P13,P14,P15,P16,P18,P20,P21,P22,P23,P25,P26,P27,P28,P29,P30, P31,P32,P33,P35,P36,P37,P38,P39,P40,P41,P42,P43,P44,P45,P46,P47, P48,P49,P50,P51,P52,P53,P54,P55,P56,P57,P58,P59,P60
Clustering	P3,P8,P11,P16,P28
Regression	P13,P21,P37,P56

ISSN: 2408-7920

Arca Academic Publisher

GBPA
Ghana Book Publishers Association

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

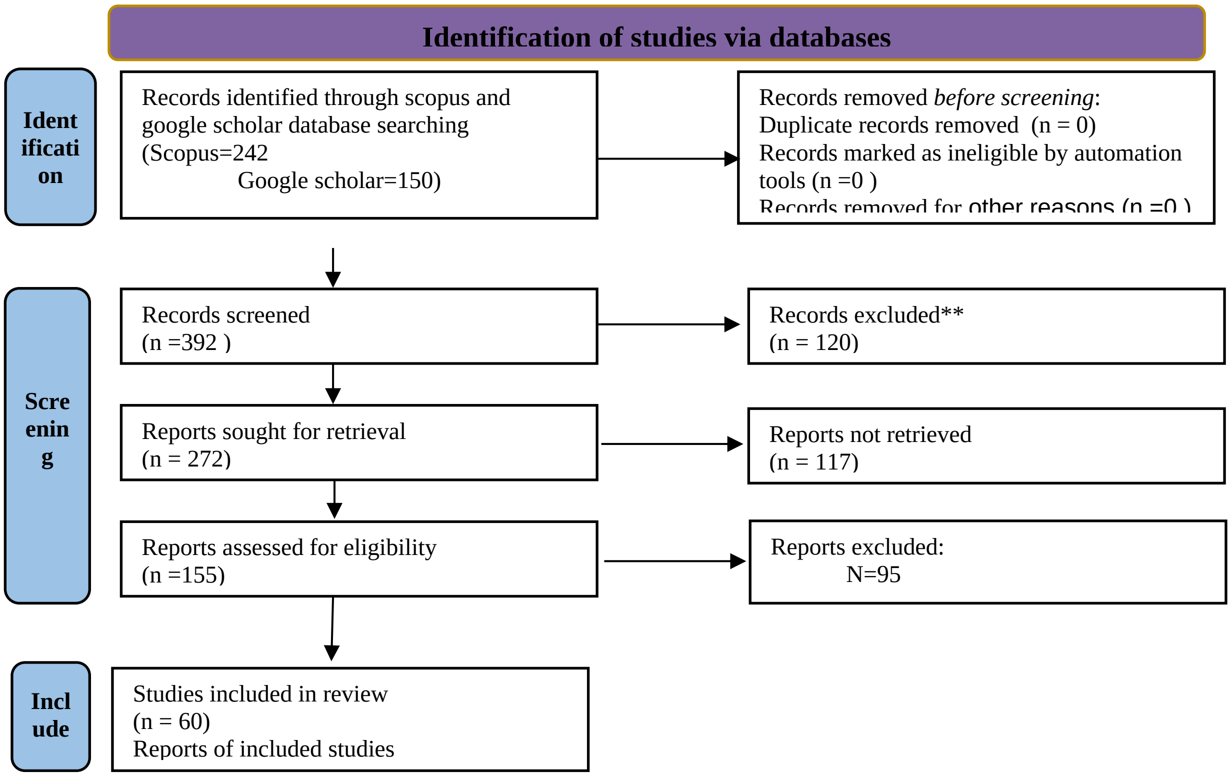

Figure 1: Standard PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram

RESULTS AND DISCUSSION

Reporting is the third stage of SLR. This part summarises and reports the analysis of the articles chosen for the study questions.

RQ1: What are the publication trends of machine learning methods in predicting and improving academic performances in HEIs?

This study analysed the research trend using four aspects: research focus, publication year, dataset origin, and study context. Figure 2 shows that 15 (25%) of the research articles focused on attributes that influenced students' performance in designing their model, 37 (62%) used data mining performance to predict academic performance, and 8 (13%) focused on the e-learning environment to design their model. Table 1 lists PS papers associated with this research focus.

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

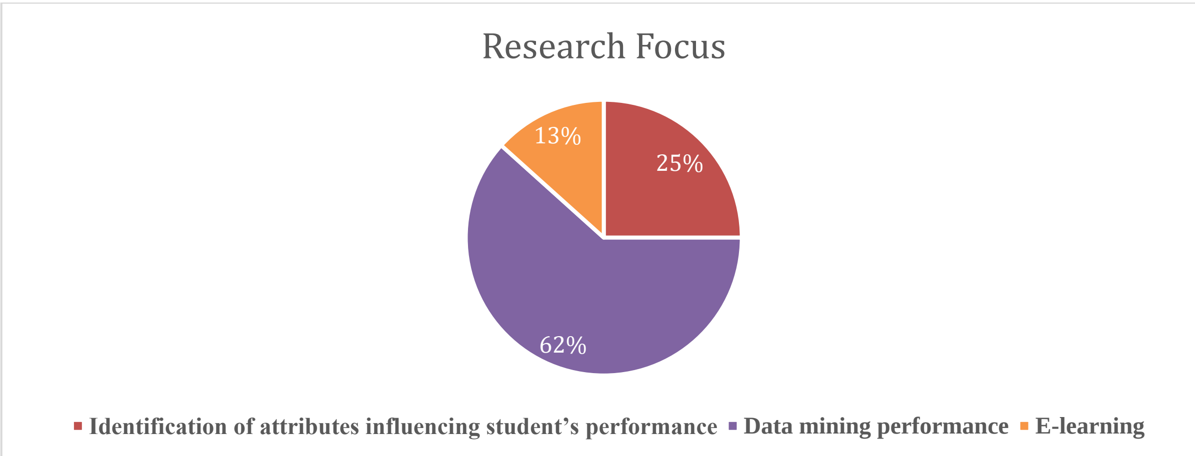

Figure 2: Research Focus

Publication year.
Figure 3a depicts the number of articles published within the chosen period. The line graph from 2018 with 6 publications shows an upward trend, which indicates studies in this context are on an upward trend. From 2019 to 2020, there was a sharp decline to 6 articles before it rose to the highest number of articles, 18 of which were published in 2021. In the following year, there was a modest decline in the number of articles published, 15 articles were published. 11 articles have been published as of August 31, 2023. Before the end of the year 2023, the number of publications will increase. Table 4 presents all PS papers associated with publication years.

Table 4: Publication year and number of selected articles published

Publication Year	Paper ID
2018	P33,P34,P43,P47,P49,P58
2019	P26,P28,P29,P30,P35,P36,P37,P39,P48,P50,P51,P57
2020	P31,P40,P42,P43,P46,P55
2021	P1,P4,P5,P9,P17,P22,P27,P32,P33,P36,P38,P41,P44,P53,P54,P56,P59,P60
2022	P6,P10,P11,P12,P13,P16,P19,P21,P23,P24,P35,P42,P45,P47,P52
2023	P2,P3,P7,P8,P14,P15,P18,P20,P25,P39,P57

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

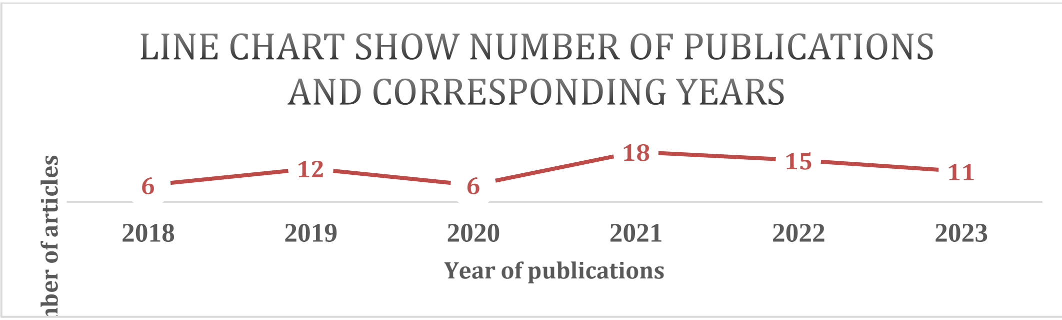

Figure 3a: Publication year of the primary papers (2018 to August 31, 2023)

Distribution by country, based on the origin of the corresponding author

This systematic evaluation determined that the primary studies (60 papers) are biassed by country. India had the most corresponding authors, with eight (13.33%), followed by China with six (10 %), Pakistan with five (8.33%), Nigeria, Bangladesh, Egypt, Saudi Arabia, Malaysia, and Australia, all of which had three (5%), and the United Kingdom (UK), Turkey, the United Arab Emirates (UAE), Thailand, and Spain, all of which had two (3.33%). The remaining thirteen countries have only one author (1.66%). Figure 3b summarises the geographical distribution of publications.

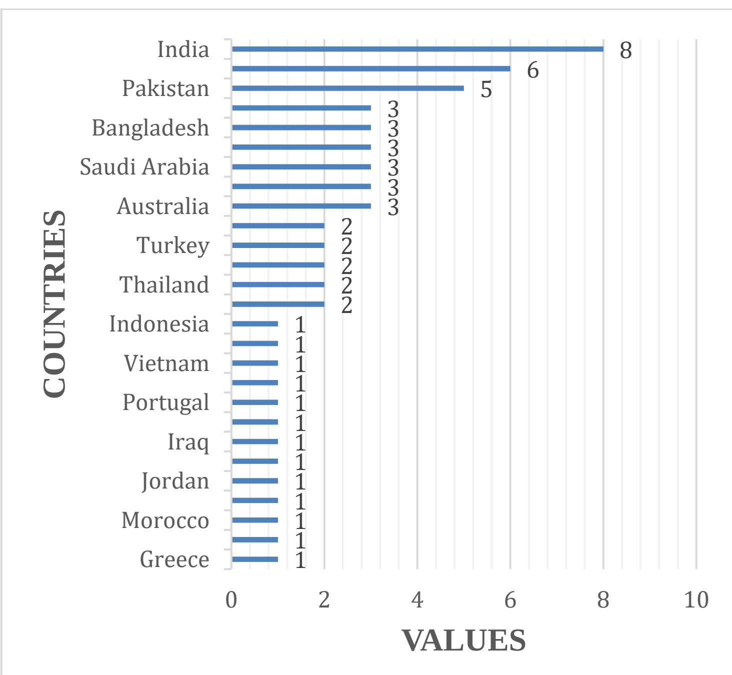

Figure 3b: Distribution by country, based on the origin of the corresponding author.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

Dataset's continent of origin.

Figure 4 shows the continents of all datasets. The dataset's countries of origin were taken and subsequently grouped into continents on the data extraction form. The highest number of articles is 25 from Asia. There were 14 articles indicated as Unknown, which indicates that the source or origin of the dataset was not mentioned in the article. The Middle East follows closely with nine articles. Not many articles were published by European countries, which is 8, while Africa follows closely with 6. America, with just two articles, met our inclusion criteria. Table 5 presents PS papers associated with the continents of the first author.

Table 5: Continental demographics of the dataset

Continent	Paper ID
Asia	P2,P3,P7,P11,P12,P13,P15,P25,P26,P28,P29,P30,P31,P33,P39,P40,P42,P43,P46,P47,P51,P53,P55 ,P57,P59
Africa	P24,P49,P52,P56,P58,P60
Middle East	P8,P19,P21,P27,P35,P36,P38,P44,P50
Europe	P1,P4,P6,P10,P34,P39,P42,P45,
America	P5,P43
Unknown	P9,P14,P16,P17,P18,P20,P22,P23,P31,P32,P37,P41,P48,P54

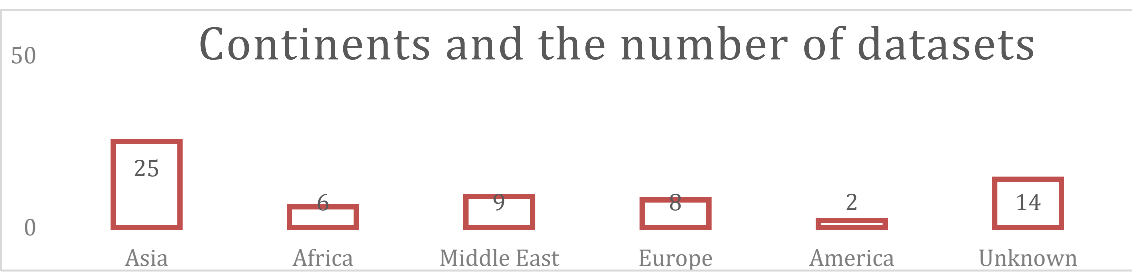

Figure 4: Continents of datasets of the primary papers

Publication outlets from the primary papers.

From Figure 5a, the primary studies (60 papers) were likewise biased towards publishing firms/outlets. The Multidisciplinary Digital Publishing Institute (MDPI) had the most papers with 7 (11.67%), followed by the Institute of Electrical and Electronics Engineers Inc (IEEE Inc) and Elsevier Ltd each with six papers (10%), Emerald Group Holdings Ltd had four papers (6.67%), Springer and International

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

Association of Online Engineering and Elsevier B.V. each with three papers (5.00%), Springer Science and Business Media Deutschland GmbH. The remaining publishing businesses each have only one paper.

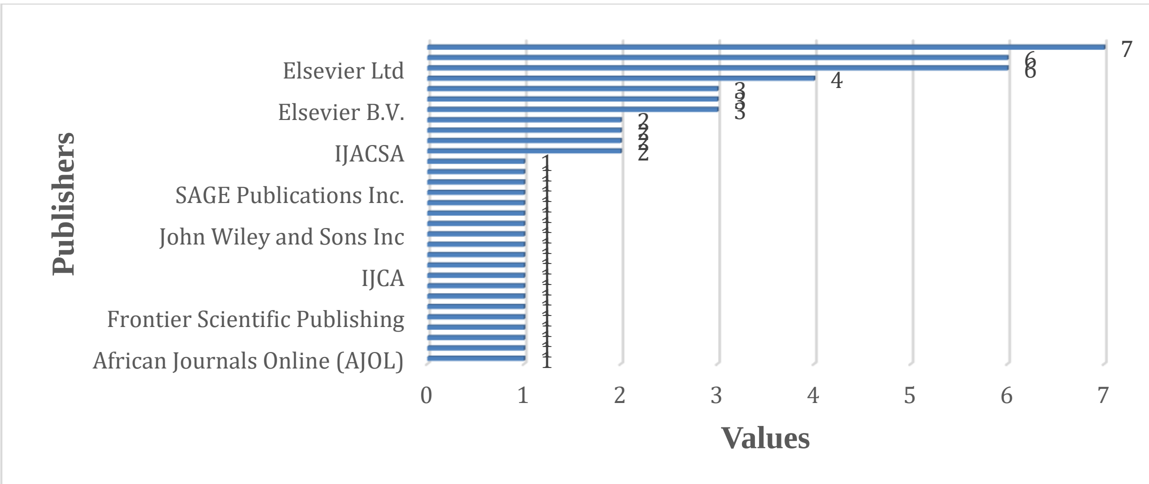

Figure 5a: Publication outlets from the primary studies.

Study context.
Higher education institutions, such as colleges and universities, are the primary focus of the study. In this context, all 60 items were centred.

RQ2: What features dominate the dataset used in predicting the academic performances of students in HEIs

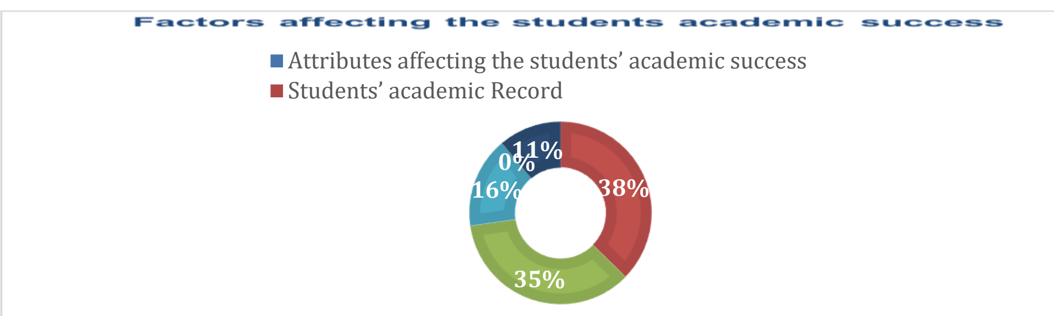

Figure 5b: Factors that affect students' academic performance

Figure 5b groups the factors that affect students' academic performance into percentages. 35% of the primary studies representing 35 articles considered students' demographic data in predicting students' academic performance. According to our research, this can be considered the highest factor affecting

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

students' academic performance. Table 4 describes each factor and the attributes it may represent. 38 %, which is the highest factor affecting student academic performance prediction, is the student's academic records of both previous education and present, and this represents 37 articles. 16 per cent of the 60 articles representing 16 in terms of number of articles took a critical look at students' activities, such as attendance to class etc, in predicting the student's performance. The student's socio-economic background is critical in every student's life, representing 11%, which is 11 articles that took a key interest in this aspect of the student. None of the included articles took interest in the student's motivation, and hence, it represents 0%.

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

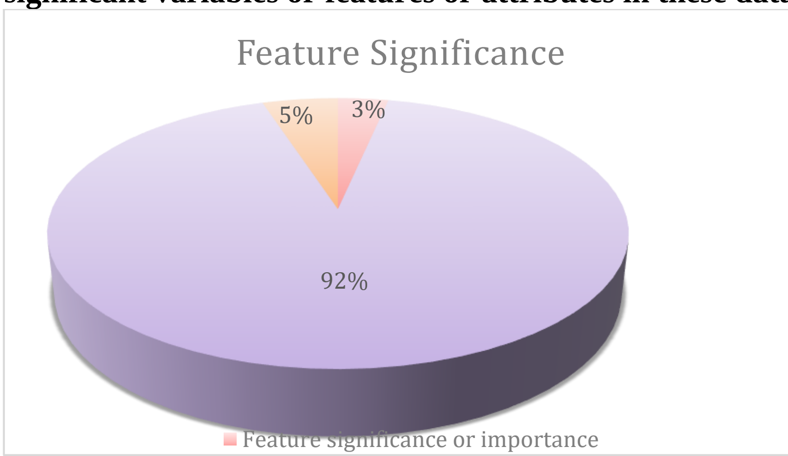

Figure 6: Articles with feature significance or importance

Feature significance is the contribution of each attribute to the model's performance. HEIs with limited resources are interested in attributes or factors that contribute primarily to a student's performance to channel resources to that area, hence the need. Figure 6 shows that 2 (3%) articles, PS 19 and PS 24, were able to identify significant features or features that contribute most to their model performance as against 92%, representing 55 articles that did not see the need for feature significance and reported on only model performances. 3 (5%) articles performed feature selection to identify correlated features to avoid multicollinearity in their features or datasets, which is suitable for the model's performance; however, that does not determine the significant features in the model performance. Students' examination scores and class participation are substantial features in PS19 and PS24. Table 6 presents PS papers that identify essential features in the model performance.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

Table 6: Articles that identify significant features in the model performance

Feature Significance to determine the most influential feature or attribute	Article ID
Significant Features identified in the modelling	P19, P24
No significant features were identified in the modelling: only reported on the modelling performance.	P1, P2, P3, P4, P5, P6,P7,P8,P9,P10,P11,P12,P13,P14,P15, P16,P18,P20,P21,P22,P23,P25,P26,P27,P28,P29,P30,P31,P32,P33,P35,P36,P37,P38,P39,P40,P41,P42,P43,P44,P45,P46,P47,P48,P49,P50,P51,P52,P53,P54,P55,P56,P57,P58, P59,P60
Feature selection was done to reduce the number of correlated features	P17,P32,P34

RQ4: What machine learning methods are used in predicting the academic performance of students in HEIs

The following are the machine learning techniques that have been used to predict student academic performances in HEIs in the sampled articles: DT-Quest algorithm, KNN, SVM, RF, ProSAP framework (collaborative data processing module for enhancing the data quality, scalable metadata clustering module for alleviating the imbalance of academic features, and XGBoost-enhanced SAP prediction module for academic performance forecasting), Artificial Neural Network (Backpropagation), NBTree, Fuzzy C-means (FCM), MLP (Multilayer Perceptron), FCM – LR, FCM – RF, Naive Bayes, One Rule, J48, JRip three, ensemble methods (Bagging, Boosting, Voting), Multiple Linear Regression Analysis (MLRA), ANN, Logistic Regression, convolutional neural network, Cobweb, Hierarchical and EM, PART, AdaBoost (ADB), Bayesian Network (BN), Naïve Bayesian (NB), Linear Regression (LR), Xgboost model, Light gbm model, Adboost model, Gdbt model, Bagging Classifier, Extra Trees, deep learning (1D-CNN, LSTM) methods BR, and AdaBoost regressor Models, Voting regression model, Hybrid 2D CNN Model, MPCA, Bayes Net-D, Native, Linear Discriminant Analysis, Logistic Model Trees (LMT), Random Tree, LADTree, Bayes Net with ADTree, Speed-constrained Multi-objective Particle Swarm Optimization (SMPSO), Sequential Minimum Optimization (SMO), REPTree, Gradient Boosting, ID3, C4.5, Simple CART, CHAID, C 5.0 and, Quick Unbiased Efficient Statistical Tree (QUEST), Passive Aggressive Classifier (PAC), Linear Discriminant Analysis (LDA), Radius Neighbours Classifier (RNC), Extra Tree (ET).

Figure 7 shows a bar chart representing the most popular machine-learning algorithm in the study. The decision tree (DT) had 32 articles that used it to predict students' academic performance in HEIs, showing its dominance. Random Forest (RF) was followed by 26 articles using it for single use. Artificial Neural Networks are also gradually gaining popularity in this area due to their ability to handle large

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

volumes of data properly; 23 primary papers used ANN. Support Vector Machine had 20 articles using it, KNN was used by 13 articles, and other machine learning techniques such as linear regression, logistic regression, Quick Unbiased Efficient Statistical Tree, etc., were used by 27 articles. Ensemble techniques such as voting, XGBoost, etc., which combine two or other machine learning techniques to enhance performance, are also well explored with 20 articles. Table 7 lists PS papers and ML algorithms used.

Table 7: Machine Learning techniques and articles

Machine learning algorithms	Articles
Decision tree	P1,P5,P7,P9,P10,P13,P14,P16,P17,P18,P22,P24,P28,P29,P32,P33,P34,P35,P36,P40,P42,P43,P44,P47,P48,P50,P51,P52,P53,P54,P57,P59
Random forest	P2,P5,P7,P8,P9,P13,P14,P15,P16,P17,P18,P19,P20,P21,P24,P29,P30,P31,P33,P36, P38,P41,P42,P43P45,P50
SVM	P2,P7,P9,P13,P14,P18,P22,P24,P25,P28,P29,P31,P33,P35,P36,P45,P47,P48,P55, P57
KNN	P2,P11,P13,P17,P18,P20,P22,P24,P28,P29,P40,P45,P59
ANN	P6,P7,P8,P9,P10,P11,P16,P20,P23,P26,P28,P31,P34,P35,P36,P42,P44,P46,P47,P48,P57,P58,P60
Naïve Bayes	P9,P13,P14,P24,P27,P28,P31,P33,P36,P38,P40,P42,P45,P47,P57,P59
Ensemble Techniques	P2,P3,P4,P9,P13,P15,P16,P17,P18,P21,P28,P29,P33,P34,P39, P40,P42,P44,P47,P50
Others	P7,P8,P10,P12,P14,P16,P20,P21,P22,P25,P26,P27,P29,P30,P31,P33,P36,P37,P38, P39,P43P45,P49,P52,P55,P56,P57,P59

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

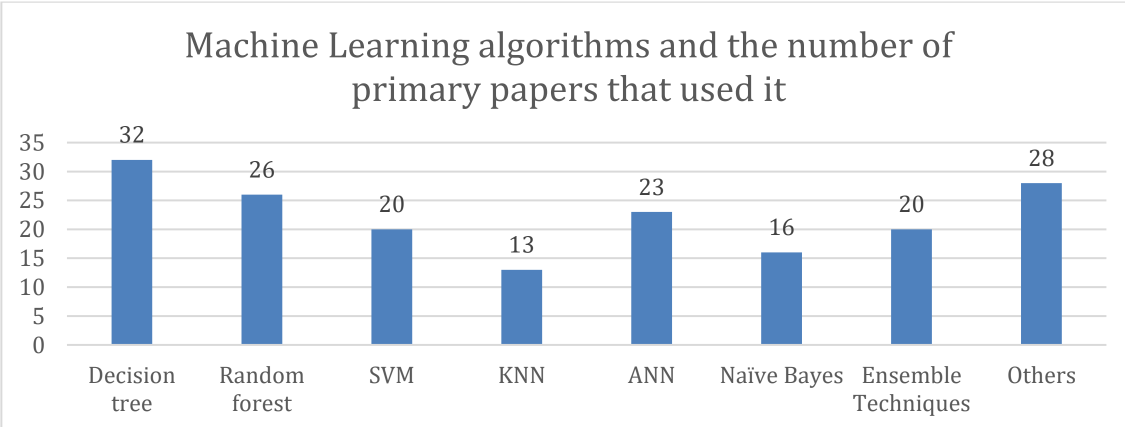

Figure 7: Machine learning algorithms and primary articles

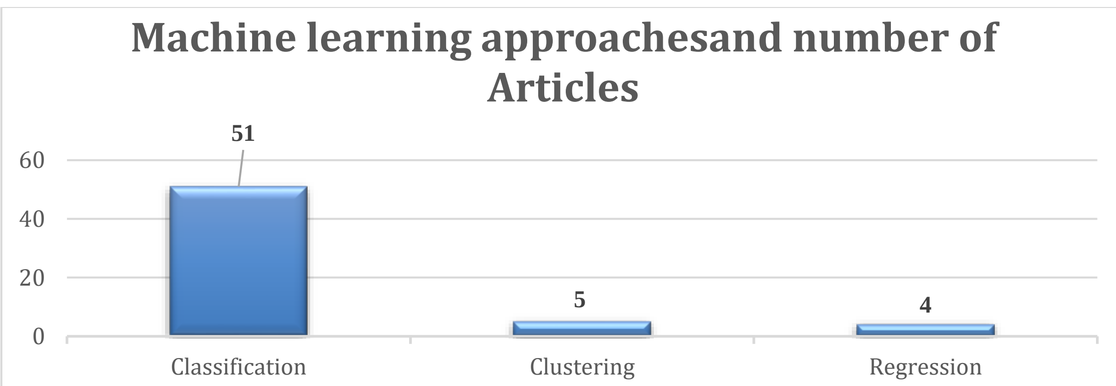

Figure 8: Machine learning approach and the number of articles

The ML predicting approaches can be categorised into three: classification, clustering, and regression. In this Systematic Literature Review, the three approaches were observed in the articles. From Figure 8, 51 articles used the classification approach, 5 articles used the clustering approach, and finally, 4 articles used the regression approach. Therefore, we can say that the classification approach of ML is the most used in student performance prediction in HEIs.

RQ5: **What gaps exist in the current literature**

Analysis from Table 2, which is the data extraction form, reveals some gaps in knowledge in the primary article. Most popular is the inability of the researchers to carry out a feature significance analysis or conduct further analysis to determine how much each feature contributes to the model performance. This is what the policymakers or management of HEIs are interested in so that they know which features to

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

address first when deciding to improve academic performance. Also, the size of the dataset used by the article is not big enough, which could affect the model's performance.

Analysis of graphic maps using VOSviewer

VOSviewer is a popular software tool for visualising and analysing bibliometric networks, primarily based on co-authorship and co-citation data. It allows researchers to create interactive visualisations of research networks, including maps of authors, documents, and keywords, to gain insights into the structure and evolution of scientific literature. VOSviewer is commonly used in bibliometrics to identify trends, research clusters, and influential papers or authors.

Co-occurrence of author keywords Figure 9a. VOSviewer (https:// www.VOSviewer.com/) was used to undertake a co-occurrence map for author keywords. We aimed to get an overview of the keywords authors used, and we hoped to identify subtopics. The terms higher education institutes (or synonyms), machine learning (or similar terms), prediction (or similar words), and academic performances (or similar terms) had to appear in the title, abstract, or keywords. VOSviewer identified one hundred eighty-two keywords, and we selected a threshold of 6 occurrences, which left 7 keywords. Figure 9a map had 3 clusters, each colour indicating a cluster, with 15 links and 32 total link strength (regarded as the weight attribute of the item). Cluster 1 consists of machine learning, student performance prediction and

education data mining terms, Cluster 2 consists of higher education and academic performance terms, and Cluster 3 has prediction and student performance terms. The term's appearance in the publications is indicated by the size of the circle given to it. Greater circles indicate more articles that contain these keywords than those with smaller circles. These terms can also be seen as the most significant in the generated co-occurrence network and as prominent themes in the research field because an item (in this case, a keyword) with a greater weight is seen as more important than an item with a lower weight. From Figure 9a, Machine learning made 20 occurrences with six links and 18 total link strengths, and educational data mining made 21 occurrences with six links and 12 total link strengths.

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

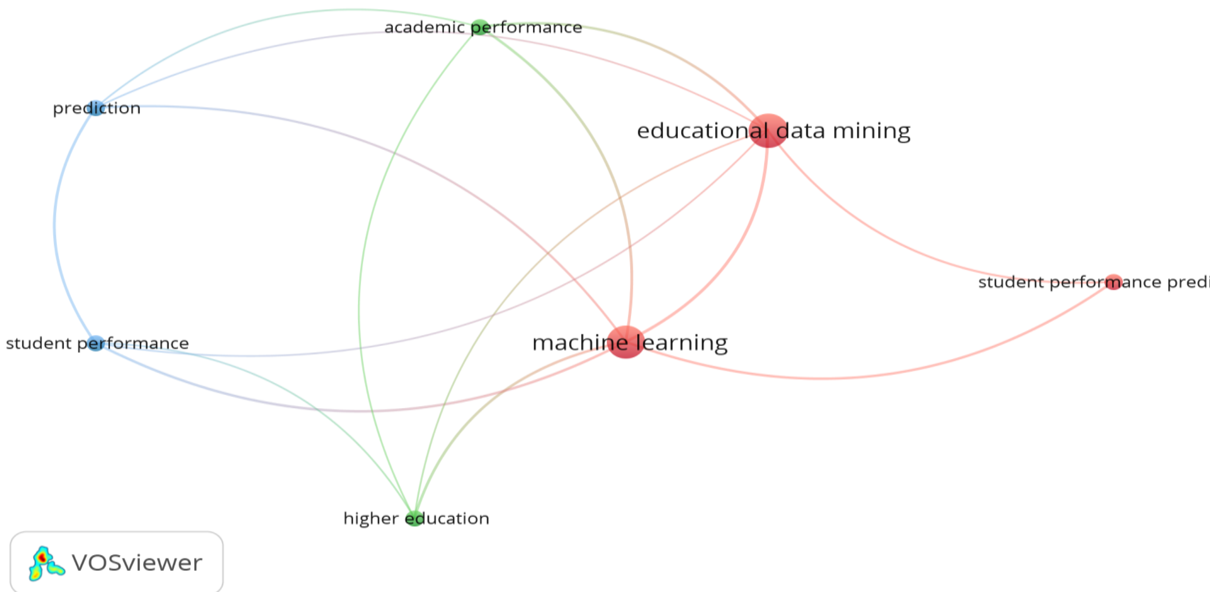

Figure 9a. Co-occurrence of author keywords with overlay visualisation.

Co-authorship by author

Figure 9b illustrates how the co-authorship network graph was constructed with the VOSviewer programme. Using the author as the unit of study, the association strength network methodology was utilised to build the network graph based on the entire counting method. Articles with more than 25 authors were not included in the analysis. Two hundred authors fulfil the criterion, with some not connected to the network. The minimum number of documents an author could have been set at 1. Six pieces make up the largest group of related items.

A co-author network created with different nodes and linkages in a single colour is displayed in Figure 9b. A node in the circle represents each author, while collaboration amongst authors is symbolised by the links that connect them. The number of publications for each author is reflected in the size of the node. The two authors' collaborative relationship is shown in the thickness of the links. One cluster of authors working together is represented by one colour. Each author in this instance has five links to a single document.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

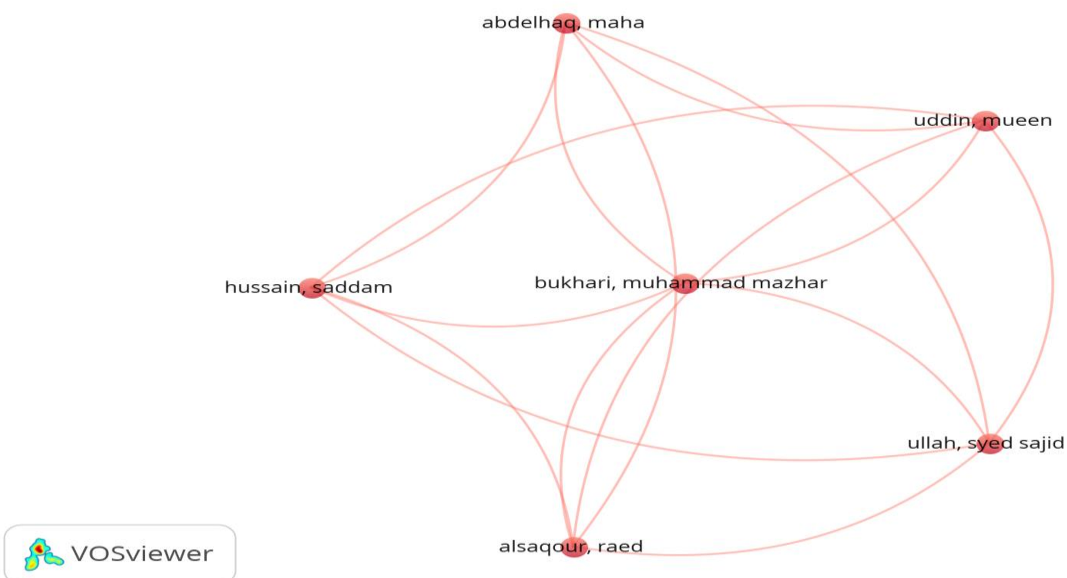

Figure 9b. Co-authorship by author with overlay visualization

Visualisation of Author Keywords

Figure 10 depicts how often the keywords appeared in the selection of primary papers. The larger the term in the picture or word cloud, the more popular it was among the author's keywords. This form of visualisation helps evaluators conduct exploratory textual analysis by detecting phrases that appear frequently in a group of interviews or documents. In the case of this study, higher education institutes, academic performance and machine learning appeared bigger than the rest which indicates the primary papers are centred around words in pictures.

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

Figure 10. Visualisation of authors' keywords

CONCLUSION

This study examined five research questions to understand better the prediction of students' academic performance using machine learning studies reported in 60 journal articles from 2018 to August 2023. The SLR reveals three categories of the research focus of these 60 studies, which include the identifying aspects influencing student performance (62%), ML algorithms performance (25%), and ML related to e-learning systems (13%). The selected journal articles are mainly about higher education institutions, and most journal articles were published in 2021, with most authors from Asia.

The SLR also reveals student academic records and demographics as popular features to predict student performance. Classification is the most used ML approach, with Decision Tree classifiers identified as the most employed algorithm, followed by Random Forest and ensemble techniques such as XGBoost, voting, etc. The datasets used for the predictions are generally small and should be critically examined in future work. Feature significance, which cannot be overlooked, should be included in modelling student academic performance since educators need to channel resources into an area that will improve performance, hence the need for feature significance in future research.

These findings can help future researchers and educators uncover potential and relevant student performance predictors in their contexts and classifier algorithms for conducting such a prediction. A good student performance prediction also allows for early intervention to improve their performance. Researchers should explore the development of machine learning models that can anticipate

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

achievements and provide customised interventions or learning paths tailored to each student's strengths and weaknesses. It would be beneficial to examine how qualitative data, such as student comments and survey responses, can be integrated into models to gain insights into challenges that quantitative data alone may overlook.

Additionally, conducting studies that monitor students' progress over time would help assess how prediction models adapt to evolving circumstances and life events. These research opportunities showcase the increasing use of analytics in education and how machine learning can help improve student support, enhance educational outcomes and facilitate data-driven decision-making in academic institutions.

REFERENCE

Abou Naaj, M., Mehdi, R., Mohamed, E. A., & Nachouki, M. (2023). Analysis of the Factors Affecting Student Performance Using a Neuro-Fuzzy Approach. *Education Sciences*, *13*(3). https://doi.org/10.3390/educsci13030313

Addo, K. D., & Sackey, S. M. (2022). DEMAND FORECAST MODELLING OF VEHICLES AS A DECISION SUPPORT: THE CASE OF TOYOTA GHANA. *African Journal of Applied Research*, *8*(2), 30–47. https://doi.org/10.26437/ajar.31.10.2022.03

Afolagboye, L. O., Ajayi, D. E., & Afolabi, I. O. (2023). Machine learning models for predicting unconfined compressive strength: A case study for Precambrian basement complex rocks from Ado-Ekiti, Southwestern Nigeria. *Scientific African*, *20*. https://doi.org/10.1016/j.sciaf.2023.e01715

Ahmed, D. M., Abdulazeez, A. M., Zeebaree, D. Q., & Ahmed, F. Y. H. (2021). Predicting University's Students Performance Based on Machine Learning Techniques. *2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 - Proceedings*, 276–281. https://doi.org/10.1109/I2CACIS52118.2021.9495862

Aliyu Sani, B., & Haruna, B. (n.d.). *Machine Learning Algorithms To Predict Student's Academic Performance*.

Alruwais, N., & Zakariah, M. (2023). Evaluating Student Knowledge Assessment Using Machine Learning Techniques. *Sustainability (Switzerland)*, *15*(7). https://doi.org/10.3390/su15076229

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

Cielen, D., Meysman, A. D. B., & Ali, M. (n.d.). *Introducing Data Science*.

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

Duong, H. T. H., Tran, L. T. M., To, H. Q., & Van Nguyen, K. (2023). Academic performance warning system based on data driven for higher education. *Neural Computing and Applications, 35*(8), 5819–5837. https://doi.org/10.1007/s00521-022-07997-6

Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. *Decision Analytics Journal*, *7*(February), 100204. https://doi.org/10.1016/j.dajour.2023.100204

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

Kitchenham, B. (2021). Kitchenham , B .: Guidelines for performing Systematic Literature Reviews in software engineering . EBSE Technical Report EBSE-2007-01 Guidelines for performing Systematic Literature Reviews in Software Engineering. *Icse*, *January 2007*, 1–57.

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

Mohd Talib, N. I., Abd Majid, N. A., & Sahran, S. (2023). Identification of Student Behavioral Patterns in Higher Education Using K-Means Clustering and Support Vector Machine. *Applied Sciences (Switzerland)*, *13*(5). https://doi.org/10.3390/app13053267

Ojokoh, B. A., Samuel, O. W., Omisore, O. M., Sarumi, O. A., Idowu, P. A., Chimusa, E. R., Darwish, A., Adekoya, A. F., & Katsriku, F. A. (2020). Big data, analytics and artificial intelligence for sustainability. In *Scientific African* (Vol. 9). Elsevier B.V. https://doi.org/10.1016/j.sciaf.2020.e00551

Roslan, M. H. bin, & Chen, C. J. (2022). Educational Data Mining for Student Performance Prediction: A Systematic Literature Review (2015-2021). *International Journal of Emerging Technologies in Learning*, *17*(5), 147–179. https://doi.org/10.3991/ijet.v17i05.27685

Russell, R. (2018). *Machine Learning: Step-by-Step Guide To Implement Machine Learning Algorithms with Python*.

Sackey, S. M., & Ngewana, T. Z. (2022). Concept Of Equivalent Stress Criterion For Predicting Failure Of Brittle Materials. *African Journal of Applied Research*, *8*(2), 332–343. https://doi.org/10.26437/ajar.31.10.2022.23

Sekeroglu, B., Abiyev, R., Ilhan, A., Arslan, M., & Idoko, J. B. (2021). Systematic literature review on machine learning and student performance prediction: Critical gaps and possible remedies. *Applied Sciences (Switzerland)*, *11*(22). https://doi.org/10.3390/app112210907

Xue, H., & Niu, Y. (2023). Multi-Output Based Hybrid Integrated Models for Student Performance Prediction. *Applied Sciences (Switzerland)*, *13*(9). https://doi.org/10.3390/app13095384

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

Appendix

List of Primary Study Papers (PS) in this Literature

PS 1: Matzavela, V., & Alepis, E. (2021). Decision tree learning through a Predictive Model for Student Academic Performance in Intelligent M-Learning environments. *Computers and Education: Artificial Intelligence*, *2*. https://doi.org/10.1016/j.caeai.2021.100035

PS2: Priyambada, S. A., Usagawa, T., & ER, M. (2023). Two-layer ensemble prediction of students' performance using learning behavior and domain knowledge. *Computers and Education: Artificial Intelligence*, *5*. https://doi.org/10.1016/j.caeai.2023.100149

PS3 : Wang, X., Zhao, Y., Li, C., & Ren, P. (2023). ProbSAP: A comprehensive and high-performance system for student academic performance prediction. *Pattern Recognition*, *137*. https://doi.org/10.1016/j.patcog.2023.109309

PS4: Chango, W., Cerezo, R., & Romero, C. (2021). Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. *Computers and Electrical Engineering*, *89*. https://doi.org/10.1016/j.compeleceng.2020.106908

PS5: Fahd, K., Miah, S. J., & Ahmed, K. (2021). Predicting student performance in a blended learning environment using learning management system interaction data. *Applied Computing and Informatics*. https://doi.org/10.1108/ACI-06-2021-0150

PS 6: Bukhari, M. M., Ullah, S. S., Uddin, M., Hussain, S., Abdelhaq, M., & Alsaqour, R. (2022). An Intelligent Model for Predicting the Students' Performance with Backpropagation Neural Network Algorithm Using Regularization Approach. *Human-Centric Computing and Information Sciences*, *12*. https://doi.org/10.22967/HCIS.2022.12.044

PS7: 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

PS8: Baig, M. A., Shaikh, S. A., Khatri, K. K., Shaikh, M. A., Khan, M. Z., & Rauf, M. A. (2023). Prediction of Students Performance Level Using Integrated Approach of ML Algorithms. *International Journal of Emerging Technologies in Learning*, *18*(1), 216–234. https://doi.org/10.3991/ijet.v18i01.35339

PS9: Evangelista, E. D. (2021). A Hybrid Machine Learning Framework for Predicting Students' Performance in Virtual Learning Environment. *International Journal of Emerging Technologies in Learning*, *16*(24), 255–272. https://doi.org/10.3991/ijet.v16i24.26151

PS10: Altun, M., Kayikçi, K., & Irmak, S. (2022). A Model Proposal for Predicting Students' Academic Performances Based on Data Mining*. *Hacettepe Egitim Dergisi*, *37*(3), 1080–1098. https://doi.org/10.16986/HUJE.2021068491

PS11: Feng, G., Fan, M., & Chen, Y. (2022). Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining. *IEEE Access*, *10*, 19558–19571. https://doi.org/10.1109/ACCESS.2022.3151652

PS12: Trakunphutthirak, R., & Lee, V. C. S. (2022). Application of Educational Data Mining Approach for Student Academic Performance Prediction Using Progressive Temporal Data. *Journal of Educational Computing Research*, *60*(3), 742–776. https://doi.org/10.1177/07356331211048777

PS13: Khamis, S., Ahmad, M., Ahmad, A., & Ahmad, M. N. (2022). Internet use behaviour model for predicting students' performance. *Expert Systems*, *39*(8). https://doi.org/10.1111/exsy.12999

PS14: Latif, G., Abdelhamid, S. E., Fawagreh, K. S., Brahim, G. Ben, & Alghazo, R. (2023). Machine Learning in Higher Education: Students' Performance Assessment Considering Online Activity Logs. *IEEE Access*, *11*, 69586–69600. https://doi.org/10.1109/ACCESS.2023.3287972

PS15: Xue, H., & Niu, Y. (2023). Multi-Output Based Hybrid Integrated Models for Student Performance Prediction. *Applied Sciences (Switzerland)*, *13*(9). https://doi.org/10.3390/app13095384

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

PS16: Yacoub, M. F., Amin Maghawry, H., Helal, N. A., Gharib, T. F., & Ventura, S. (n.d.). An Enhanced Predictive Approach for Students' Performance. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 13, Issue 4). www.ijacsa.thesai.org

PS17: Sahlaoui, H., Alaoui, E. A. A., Nayyar, A., Agoujil, S., & Jaber, M. M. (2021). Predicting and Interpreting Student Performance Using Ensemble Models and Shapley Additive Explanations. *IEEE Access*, *9*, 152688–152703. https://doi.org/10.1109/ACCESS.2021.3124270

PS18: 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

PS19: Nachouki, M., & Naaj, M. A. (2022). Predicting Student Performance to Improve Academic Advising Using the Random Forest Algorithm. *International Journal of Distance Education Technologies*, *20*(1). https://doi.org/10.4018/IJDET.296702

PS20: 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

PS21: Abdullah, M., Al-Ayyoub, M., AlRawashdeh, S., & Shatnawi, F. (2023). E-learningDJUST: E-learning dataset from Jordan university of science and technology toward investigating the impact of COVID-19 pandemic on education. *Neural Computing and Applications*, *35*(16), 11481–11495. https://doi.org/10.1007/s00521-021-06712-1

PS22: Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks. *IEEE Access*, *9*, 140731–140746. https://doi.org/10.1109/ACCESS.2021.3119596

PS23: 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

PS24: El-Keiey, S., Elmenshawy, D., & Hassanein, E. (n.d.). Student's Performance Prediction based on Personality Traits and Intelligence Quotient using Machine Learning. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 13, Issue 9). www.ijacsa.thesai.org

PS25: Sabri, M. Z. M., Majid, N. A. A., Hanawi, S. A., Talib, N. I. M., & Yatim, A. I. A. (2023). Prediction Model based on Continuous Data for Student Performance using Principal Component Analysis and Support Vector Machine. *TEM Journal*, *12*(2), 1201–1210. https://doi.org/10.18421/TEM122-66

PS26: Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, *1*(9). https://doi.org/10.1007/s42452-019-0884-7

PS27: Mohammad Abu-dalbouh, H. (2021). APPLICATION OF DECISION TREE ALGORITHM FOR PREDICTING STUDENTS' PERFORMANCE VIA ONLINE LEARNING DURING CORONAVIRUS PANDEMIC. *Journal of Theoretical and Applied Information Technology*, *15*, 19. www.jatit.org

PS28: Francis, B. K., & Babu, S. S. (2019). Predicting Academic Performance of Students Using a Hybrid Data Mining Approach. *Journal of Medical Systems*, *43*(6). https://doi.org/10.1007/s10916-019- 1295-4

PS29: 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

PS30: Aman, F., Rauf, A., Ali, R., Iqbal, F., & Khattak, A. M. (2020, July 1). A Predictive Model for Predicting Students Academic Performance. *10th International Conference on Information, Intelligence, Systems and Applications, IISA 2019*. https://doi.org/10.1109/IISA.2019.8900760

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

PS31: 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

PS32: Al Karim, M., Yeasmin Ara, M., Mahadi Masnad, M., Rasel, M., & Nandi Professor, D. (n.d.).

PS33: 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

PS34: 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

PS35: Musaddiq, M. H., Sarfraz, M. S., Shafi, N., Maqsood, R., Azam, A., & Ahmad, M. (2022). Predicting the Impact of Academic Key Factors and Spatial Behaviors on Students' Performance. *Applied Sciences (Switzerland)*, *12*(19). https://doi.org/10.3390/app121910112

PS36: Sadiq, M. H., & Ahmed, N. S. (2019). Classifying and predicting students' performance using improved decision tree C4.5 in higher education institutes. *Lubricants*, *7*(12), 1291–1306. https://doi.org/10.3844/jcssp.2019.1291.1306

PS37: Palmer, S. (n.d.). *Modelling Engineering Student Academic Performance Using Academic Analytics**.

PS38: Kamal, M., Chakrabarti, S., Ramirez-Asis, E., Asis-Lopez, M., Allauca-Castillo, W., Kumar, T., Sanchez, D. T., & Rahmani, A. W. (2022). Metaheuristics Method for Classification and Prediction of Student Performance Using Machine Learning Predictors. *Mathematical Problems in Engineering*, *2022*. https://doi.org/10.1155/2022/2581951

PS39: 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

PS40: Singh, R. (2020). Machine Learning Algorithms and Ensemble Technique to Improve Prediction of Students Performance. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(3), 3970–3976. https://doi.org/10.30534/ijatcse/2020/221932020

PS41: 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

PS42: Realinho, V., Machado, J., Baptista, L., & Martins, M. V. (2022). Predicting Student Dropout and Academic Success. *Data*, *7*(11). https://doi.org/10.3390/data7110146

PS43: Sandoval, A., Gonzalez, C., Alarcon, R., Pichara, K., & Montenegro, M. (2018). Centralized student performance prediction in large courses based on low-cost variables in an institutional context. *Internet and Higher Education*, *37*, 76–89. https://doi.org/10.1016/j.iheduc.2018.02.002

PS44: Ahmed, D. M., Abdulazeez, A. M., Zeebaree, D. Q., & Ahmed, F. Y. H. (2021). Predicting University's Students Performance Based on Machine Learning Techniques. *2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 – Proceedings*, 276–281. https://doi.org/10.1109/I2CACIS52118.2021.9495862

PS45: Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1). https://doi.org/10.1186/s40561-022-00192-z

PS46: 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).

PS47: Duong, H. T. H., Tran, L. T. M., To, H. Q., & Van Nguyen, K. (2023). Academic performance

African Journal of Applied Research
Vol. 11, No. 1 (2025), pp.536-559
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v11i1.
Received: June 18, 2024
Peer reviewed: September 4, 2024
Revised: November 25, 2024
Published: January 2025

warning system based on data driven for higher education. *Neural Computing and Applications*, *35*(8), 5819–5837. https://doi.org/10.1007/s00521-022-07997-6

PS48: Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, *98*, 166–173. https://doi.org/10.1016/j.chb.2019.04.015

PS49: Saheed, Y. K., Oladele, T. O., Akanni, A. O., & Ibrahim, W. M. (2018). Student performance prediction based on data mining classification techniques. *Nigerian Journal of Technology*, *37*(4), 1087. https://doi.org/10.4314/njt.v37i4.31

PS50: Prasanalakshmi, B., & Farouk, A. (2019). Classification and Prediction of Student Academic Performance in King Khalid University- A Machine Learning Approach. *Indian Journal of Science and Technology*, *14*(12), 1–6. https://doi.org/10.17485/ijst/2019/v12i14/142792

PS51: Gotardo, M. A. (2019). Using Decision Tree Algorithm to Predict Student Performance. *Indian Journal of Science and Technology*, *12*(8), 1–8. https://doi.org/10.17485/ijst/2019/v12i5/140987

PS52: MURAINA, I. O., Aiyegbusi, E., & Abam, S. (2023). Decision Tree Algorithm Use in Predicting Students' Academic Performance in Advanced Programming Course. *International Journal of Higher Education Pedagogies*, *3*(4), 13–23. https://doi.org/10.33422/ijhep.v3i4.274

PS53: Varade, R. V., & Thankanchan, B. (2021). Academic Performance Prediction of Undergraduate Students using Decision Tree Algorithm. *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology*, *13*(SUP 1), 97–100. https://doi.org/10.18090/samriddhi.v13is1.22

PS54: Hoque, M. I., Azad, A. kalam, Abu Hurayra Tuhin, M., & Salehin, Z. U. (2020). University students result analysis and prediction system by decision tree algorithm. *Advances in Science, Technology and Engineering Systems*, *5*(3), 115–122. https://doi.org/10.25046/aj050315

PS55: Singh, R., & Pal, S. (2020). Application of Machine Learning Algorithms to Predict Students Performance Application of Data Mining Techniques with special reference to academic Performance Monitoring and evaluation in higher education View project Designing of Transition Metal-Macrocycles For Biological Applications View project Application of Machine Learning Algorithms to Predict Students Performance. *Article in International Journal of Advanced Research in Computer Science*, *29*(5), 7249–7261.

PS56:Yakubu, M. N., & Abubakar, A. M. (2022). Applying machine learning approach to predict students' performance in higher educational institutions. *Kybernetes*, *51*(2), 916–934. https://doi.org/10.1108/K-12-2020-0865

PS57: Haerani, E., Syafria, F., Lestari, F., & Marzuki, I. (n.d.). CLASSIFICATION ACADEMIC DATA USING MACHINE LEARNING FOR DECISION MAKING PROCESS. In *Journal of Applied Engineering and Technological Science* (Vol. 4, Issue 2).

PS58: Yohannes, E., & Ahmed, S. (2018). Prediction of Student Academic Performance using Neural Network, Linear Regression and Support Vector Regression: A Case Study. *International Journal of Computer Applications*, *180*(40), 39–47. https://doi.org/10.5120/ijca2018917057

PS59: Yaacob, W. F. W., Nasir, S. A. M., Yaacob, W. F. W., & Sobri, N. M. (2019). Supervised data mining approach for predicting student performance. *Indonesian Journal of Electrical Engineering and Computer Science*, *16*(3), 1584–1592. https://doi.org/10.11591/ijeecs.v16.i3.pp1584-1592

PS60: Albarka Umar, M. (2019). Student Academic Performance Prediction using Artificial Neural Networks: A Case Study. In *International Journal of Computer Applications* (Vol. 178, Issue 48).