Decision Analytics Journal 7 (2023) 100204

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

A systematic review of the literature on machine learning application of determining the attributes influencing academic performance

Iddrisu Issah [a,*], Obed Appiah [a], Peter Appiahene [a], Fuseini Inusah [b]

[a] University of Energy and Natural Resources, Department of Computer Science and Informatics, Sunyani, Ghana
[b] University for Development Studies, Faculty of Education, Department of ICT, Tamale, Ghana

ARTICLE INFO

Keywords:
Machine learning
Data mining
Demographic
Pre-tertiary
Predictive analytics
Learning analytics

ABSTRACT

Academic institutions operate in an extremely demanding and competitive environment. Some difficulties confronting most schools are delivering high-quality education to the students, developing systems for evaluating student performance, analyzing performance, and recognizing the future demands of their learners. Also, due to the paradigm shift due to the computerization of school data management, educational stakeholders, including the machine learning (ML) community, have taken an interest in analyzing performance traits using academic and non-academic factors. This systematic literature review identifies various machine learning methods based on 84 selected publications. It shows how researchers have been able to pattern-map student characteristics and their influence on school performance. An attempt is made to answer how the overall study coverage of student characteristics and the ML methods are employed to predict students' performance. An analysis of the 84 papers highlights that, student characteristics predominantly influencing performance are academic and demographic attributes. The study further shows that classification and decision trees are the most widely used methods and algorithms. The review also reveals population and practical knowledge gaps due to a lack of research on basic academic performance and prescription of intervention plans for averting poor performance through mapping these influential characteristics to student accomplishment. To bridge these perceived gaps, the scope of the population sample needs a benchmarked dataset and embedding the appropriate intervention outlines that will map the learner's performance early in their school life.

Contents

* Corresponding author.
E-mail addresses: issah.iddrisu.stu@uenr.edu.gh (I. Issah), obed.appiah@uenr.edu.gh (O. Appiah), peter.appiahene@uenr.edu.gh (P. Appiahene), finusah@uds.edu.gh (F. Inusah).

https://doi.org/10.1016/j.dajour.2023.100204
Received 3 October 2022; Received in revised form 22 February 2023; Accepted 17 March 2023
Available online 21 March 2023

1. Introduction

Student assessment takes a center stage in educational activities when it comes to the evaluation of educational attainment of students in educational institutions in Ghana.

Students' success is a key performance indicator for first and second cycle schools as well as institutions of higher learning [1]. Improved performance enhances the school's ranking as a sign of good quality delivery. In [2], it is postulated that excellent performance at the pre-tertiary schools also gives students the green light to be admitted to the best tertiary institutions as well as afford them the opportunity to pursue their desired programs.

Nowadays, academic establishments operate in an extremely demanding and competitive settings [3]. Accordingly, some of the difficulties that most schools confront today include delivering high-quality education to their students, developing systems for evaluating student performance, analyzing performance, and recognizing future demands of their learners. In schools, student academic intervention plans have been introduced to help students cope with difficult situations in learning. [3] again reported that school administrators and education stake holders also benefit from the schools' excellent development and evolution of intervention plans, which are based on student performance predictions at the point of entry and in subsequent semesters.

To this end, players in education and the global machine learning (ML) community are much interested in predicting students' success in pre-tertiary or higher education settings [4]. To create and apply predictive models, variables that are correlated with the values to be predicted must be gathered and analyzed. According to [4], a variety of variables that affect students' performance can be grouped into different categories. These variables include those relating to the students' previous education, the use of e-learning, socio-demographic traits, data from their social media platforms, cognitive and behavioral practices, formative assessment, extracurricular and co-curricular activities, school environment, and family dynamics, among others.

The degree of interest that students have in their studies varies, according to [4], and although some students can independently study via e-learning, others may require the assistance of teachers who will tailor their teaching strategies to the requirements of the students. Others may be able to pass their tests using all the resources provided by the school or institution, while some students from more affluent backgrounds may be able to secure outside financing to pay for the expense of additional lessons needed to meet their educational needs.

As a result of technological advancements utilized to manage, measure, and retain student data, evaluating students' performance has recently become a challenging endeavor given the variables involved in the prediction process [5]. In the view of [6], the complexity and enormous amount of data stored in educational databases, the data generated by these systems sometimes overwhelms decision-makers in the field of education. But [7] noted in their study that recent technological developments have made it feasible for computationally intensive prediction methodologies, such as machine learning, to be a workable substitute for a variety of applications, including educational decision support systems (EDSS).

To this end, academics with a variety of research interests have worked to identify the factors that significantly affect learning outcomes as well as the most effective teaching strategies. These initiatives aim to support educational institution administrators in finding ways to give their students the greatest possible learning environment [8]. In this way, teachers would be able to design adaptable educational materials and feedback according to student approval to direct students' learning development.

Regardless of evidence of the existence of numerous reviews related to performance prediction, a few of them relates to the feasibility of merging diverse students' characteristics and ML methods to ensure the accuracy and precision of predictions.

Again, there is an apparent dearth of research demonstrating the influence of socio-demographic variables on student achievement. Therefore, the objectives in the review are to comprehensively map, measure, and investigate accessible articles that were released between 2016 and 2022 on machine learning methods and their implementations in student performance predictions.

2. Related reviewed literature

With the increased interest in predicting student success across educational institutions, academics have put in a concerted effort to identify the possible important variables controlling how learners perform.

The influence of a wide range of factors on students' prediction accuracy has been reviewed in the literature. The focus of this research was on prior academic accomplishments, demographic traits, student behavioral traits, psychological variables, family socioeconomic background, and school environment (as shown in Table 4.1). In this systematic literature review (SLR), 57% of the total papers on analyses was found to have used prior academic achievements and demographic characteristics in predicting the student learning outcomes. This discovery is consistent with the findings of a review of research on predicting academic achievement in higher education by [9]. They discovered that 69% of research studies used academic accomplishments and demographic characteristics as the primary contributors to the academic success of higher education students. Their review of research, on the other hand, fell short in broadening the study to encompass lower levels learners' educational traits.

[10] undertook a survey of papers on machine learning algorithms to predict academic achievement published between 2019 and 2021. Eleven papers were examined in all. The study was mostly focused on the use of data generated during a student registration, student demographics and life styles such as proficiency in task performance, and style of learning, as well as sleeping habits and exercises. The artificial neural network (ANN) was discovered to be the most frequently used machine learning algorithm. According to the 11 publications analyzed, the most important factors in determining students' success were the students' attentiveness in theory class, test results in Moodle, and engagement in Moodle discussion boards. The review could not however properly state the effects of students' demographics on their achievements as observed in literature.

Similarly, [11] did an SLR over a ten-year period (2010–2020) based on 176 articles. This review focused on machine learning and student achievement predictions, as well as the gaps and potential solutions in existing research. 62% of the research was shown to be essentially classification approaches. The review also found 76.60% of the studies were conducted using higher education data sets and 23.40% of the studies were based on basic education datasets. The

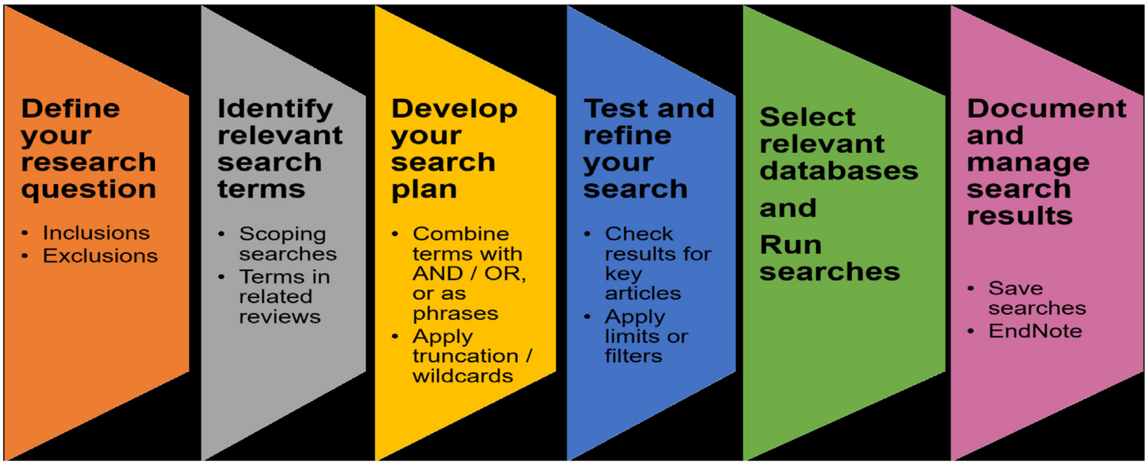

Fig. 3.1. Flowchart Depicting Okoli's guide for systematic literature review.

study discovered that students' behaviors, demographics, and social lives were the most relevant determinants in performance prediction. This study highlights the previously documented marginalization of data relating to lower educational settings.

Another review study on existing literature aimed at discovering algorithms that may be modeled to predict student performance and boost learning was done by [12]. They compared 10 distinct algorithms based on their accuracies of predictions on student performance. They found, however, that choosing the optimal algorithms for predicting student achievement was problematic considering the variety of parameters employed by different researchers throughout the literature. They also opined these socioeconomic factors: the family finances, parental education level, and work status of the parents or students as having significant impact on learner success. They discovered, however, that these variables are frequently disregarded when predicting student performance.

[13] assessed 56 publications in an SLR based on 10 study quality evaluation questions as a review guide, of which only 34 papers offered the features as well as their value in predicting student success. The characteristics addressed in the study were roughly classified as demographic, academic, and behavioral. According to [13], standardizing predictions through the use of benchmark datasets will improve the practicality of ML model selection and prediction of academic achievement among students. However, the study failed to identify a consistent approach for selecting the most relevant elements to improve performance while decreasing learner attrition.

3. Methodology

To guarantee consistency in reviewing all various studies linked to the given topic, a comprehensive SLR must always be undertaken using an impartial study approach. To do this, the three different stages: "the review plan, how to carry out the review, and review reporting process" recommended by [14], was adopted along with [15] for conducting a standalone SLR. With these two methods, a comprehensive, standardized process for systematic literature review is introduced to the study. Despite being primarily focused on educational data mining (EDM) research, the study is sufficiently detailed to be useful and applicable to academics in any social scientific discipline. The entire flowchart for Okoli's guidelines for systematic literature review is shown in Fig. 3.1.

3.1. Planning the review

A structured review was conducted to describe the most recent seven-year publications of works on success of students in learning.

According to [14], initiating every review process involve planning. Consequently, there is the need to determine and design a well-prepared plan for the review process.

Table 3.1
Criteria for drafting systematic literature review questions [14].

Criteria	Detail
Population	Pre-tertiary and higher institutions of learning
Intervention	Methods, algorithms and techniques employed in predictions
Outcome	Best performance indicators', top characteristics or factors, as well as effective prediction strategies or approaches
Context	Educational establishments and Maximum performance attainment of Students

3.1.1. Research objectives

Because the reviewer's primary aim is to determine the significance of this SLR, the following objectives are given for the review:

1. To discover the dominant characteristics employed in student performance prediction.
2. To discover approaches often adopted by scholars in predicting student achievement.
3. To ascertain which algorithms are most often employed in predicting student success.
4. To detect any gap in knowledge past studies might have over looked.

3.1.2. Research questions

Formulation of an SLR question is the most crucial step in a systematic review methodology.

Based on [14], the population, intervention, outcomes, and context (PIOC) viewpoints of the study criteria must be taken into account when developing research questions. The criteria for research questions are shown in Table 3.1.

Based on Table 3.1, the fundamental SLR query would be how to identify the overall study coverage of student characteristics as well as the appropriate ML methods when predicting students' performance?

The following particular questions were developed in response to this query

1. Which characteristics are frequently employed by academics when predicting student learning outcomes?
2. Which ML methods are frequently employed by researchers when predicting student learning outcomes?
3. Which algorithms or techniques are most effective for predicting student performance?

3.2. Conducting the review

3.2.1. Source of information and exploration approach

Credible, well-planned sources of information and search queries are essential for achieving the objectives and outcomes of this study.

Table 3.2
Publications initially sourced from four databases consulted.

Identifiers	Data bases	Date of access	URL	Count
DB1	IEEE explorer	13/10/2021	https://www.ieee.org	82
DB2	ERIC	15/11/2021	https://www.eric.ed.gov	56
DB3	ScienceDirect	11/10/2021	https://www.sciencedirect.com/	42
DB4	Springer Link	11/11/2021	https://link.springer.com/	32

A thorough and a well-organized investigation was carried out to answer the study questions. Between October 11 and December 15, 2022, searches were made in five online databases (IEEE Xplorer, ERIC, Science Direct and Springer Link). Between February 13 and 18, 2023, a follow-up search was conducted again to capture possible publications made in 2022 on the topic. However, all studies from January, 2023 onwards could not be captured since the search was conducted within the fourth quarter of the year 2022.

To ensure the capture of related articles, various search query formats, including key terms, Boolean, and wild card operators, were designed, assembled in various combinations, and executed to ensure the capture of relevant articles. Examples are;

- "Student Success Prediction" & "Educational Data Mining"
- "Students" + "Learning Curve Prediction" & "Machine Learning"
- "Students" + "Learning pattern Prediction" &"Data Mining"
- "Approaches" OR "methods" OR "techniques" & for "prediction" & "achievement of students" OR "success of students" & "Machine learning"
- Predict* & "student success" & "machine learning".

Table 3.2 contains the respective databases, the dates, and the total number of research papers retrieved for the review.

3.2.2. The review paper selection procedure

The main criteria for selection in this review was developed to determine key studies based on research papers' identification, screenings, suitability verification, and inclusion criteria satisfaction.

Studies that looked into how to predict students' performance at pre-tertiary or higher education institutions were captured as well.

Studies were chosen for the first round of screening using publication "titles and abstracts". Studies were chosen for the second stage based on analysis of the "full-text" version. When there were any ambiguities, studies were also taken into consideration for a full-text review.

The inclusion and exclusion criteria

The major purpose was to incorporate many publications relating to the research objectives as possible between 2016 and 2022 inclusive. Secondly, the guidelines were established to determine whether a paper ought to be included in the analysis or not.

Conditions for inclusion

1. Research publications based on predictions of students' academic performance
2. Research on student performance using ML or DM methods.
3. Articles published between 2016 and 2022 are eligible.
4. English language written works and other papers with English language equivalent translation were captured.
5. Submission and publication of research papers in peer-reviewed journals or conferences

Conditions for Exclusion

1. Research that does not employ ML or DM techniques
2. Papers with duplicate entries
3. Unrelated headings, abstracts, and content
4. Papers based on suggested methods or that did not conduct experiments
5. Secondary or tertiary study with the primary purpose of reporting the findings of a systematic review or mapping inquiry or other techniques while presenting a systematic review or map-based research were excluded.
6. Methodological research guidelines or handbooks with a broad focus (i.e., without the use of DM or ML) were also not captured.
7. Submissions where the only formats available were slide demonstrations or long summaries. Studies that were just given as an abstract or on slides would not have enough information to be considered for the collection of chosen studies.
8. Recommendations for performing or presenting research articles (i.e., empirical evidence evaluating a technique) rather than instructions for assessing the effectiveness of research articles. As a result, the scope of our study excludes processes for performing and reporting primary studies.

Table 3.3
Selected publication for the study by year of publication.

Year of publication	References	Paper count	
		No. of papers	% Count
2016	[16–26]	11	13.09%
2017	[27–40],	14	16.67%
2018	[41–47],	7	8.33%
2019	[48–60]	13	15.48%
2020	[61–67].	7	8.33%
2021	[68,69,1,70–76,2,77–79]	14	16.67%
2022	[80–97].	18	21.43%

Table 3.4
Distribution of publications included in the study according to databases consulted after screening.

Identifiers	Data bases	URL	Count	% Count
DB1	IEEE Xplorer	https://www.ieee.org	25	29.8%
DB2	ERIC	https://www.eric.ed.gov	11	13.1%
DB3	Science Direct	https://www.sciencedirect.com/	19	22.6%
DB4	Springer Link	https://link.springer.com/	29	34.5%

A search was conducted to compile a list of studies for further review. Mendeley was employed to keep track of the references, to exclude duplicates, and identify publication years. These bibliographies include research that perfectly meets the requirements for inclusion. The 84 papers that were left after successfully executing inclusion and exclusion criteria were arranged into groups based on the year they were published. The number of papers published per year is described in Table 3.3. The review has taken into account every paper cited in the study.

Following the use of duplication screening, 62 publications got eliminated. Using titles and abstracts screening criteria 38 publications were removed, 24 were also excluded after full text screening and 4 were eliminated based on the comprehensive full-text and content analysis. This is depicted in Fig. 3.2. The overall publications captured 84 papers. The distribution of the 84 publications among the four databases consulted is indicated in Table 3.4.

The distribution of the included publications by year is shown in Fig. 3.3. From the figure, it is realized that, after a sharp reduction in published articles from 2019 to 2020 in relation to students' performance prediction, the number of articles increased dramatically from 2021 through to 2022. This shows that numerous academics have recently become interested in using ML techniques to predict achievement of students. Fig. 3.3 demonstrates that the majority of the included publications were released in 2017 and 2021 (N = 14, 16.67%), 2019 (13, 15.48), and a further upsurge in publications in the year 2022 (18, 21.43%).

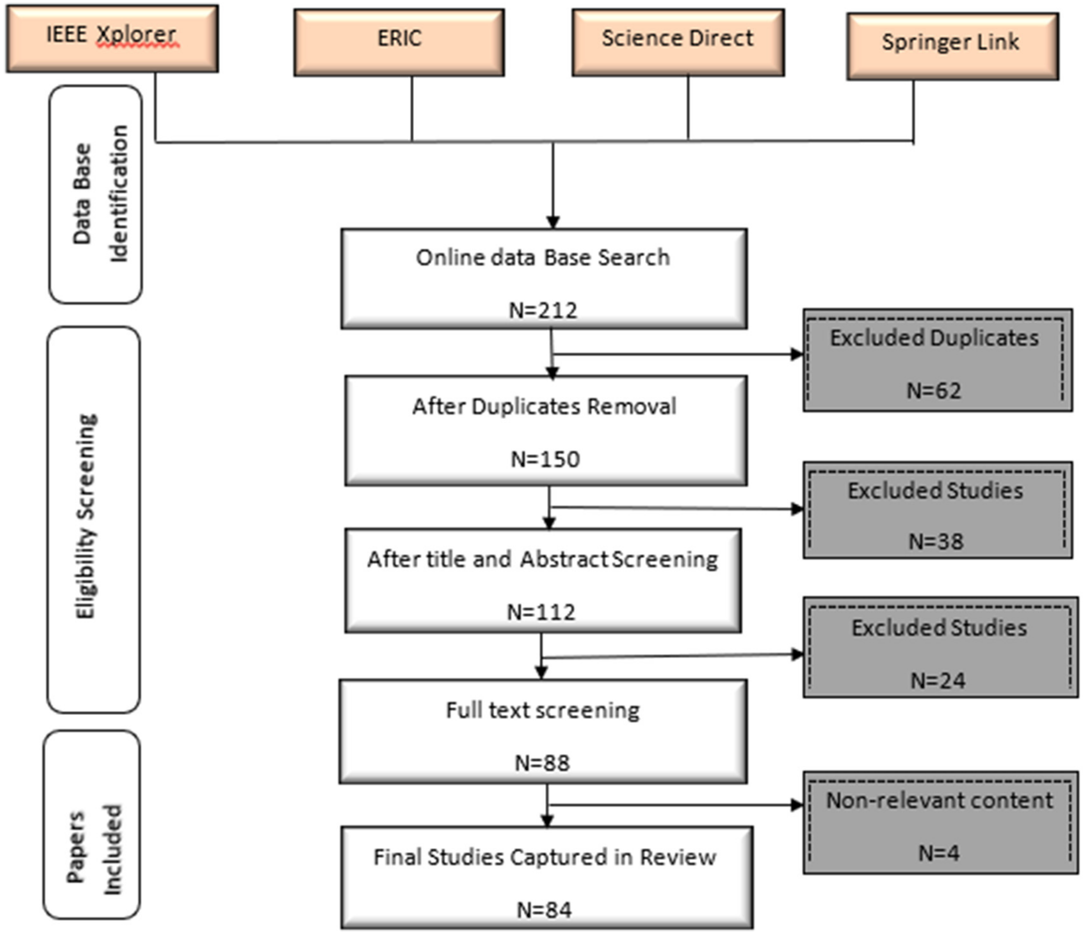

Fig. 3.2. A modified study screening and selection process adopted from [7,5].

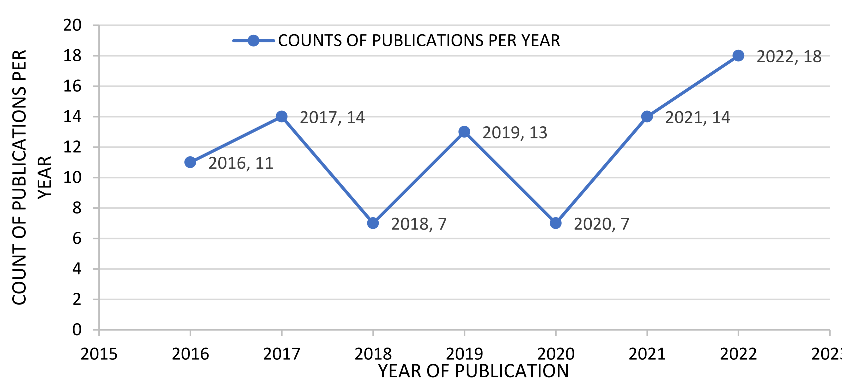

Fig. 3.3. Analysis of publications on students' performance predictions on yearly basis.

4. Results and discussion

This section examines the characteristics, techniques, and algorithms employed during this review research. In general, two important factors determine the accuracy of learning outcome of students. These include qualities/traits of data sources, as well as techniques being deployed to analyze academic related data to find trends and predictions [5]. The results analysis is grounded on the research objectives of the study.

RQ1: Which characteristics are frequently employed by academics when predicting student learning outcomes?

Several studies have revealed a variety of characteristics that influence prediction accuracy. Although having various labels, several qualities have comparable meanings and can be put into a single unit. Education history, demographics, social network information, behavioral factors, extracurricular activities, school design, academic achievement, and socioeconomic background data on their families are just a few examples.

For the purpose of this review, some variables were identified and collected. The six categories of features that were found in publications on the subject under review were academic features, demographic features, behavioral characteristics, psychological, family background, and school environment.

84 key papers were reviewed for this investigation. In 34.20% of instances, scholars used attributes from the academic performance group to predict students' achievement. Demographic characteristics

Table 4.1
Attributes frequently deployed in predicting students' achievement.

Attributes	Attribute Domain	Freq.	Reference
Academic performance	grade point average (GPA), Grade level, High school score, attendance to lessons, number of courses per semester	39	[61,16,27,41,48,28,29,51,49,30,68,31,50,20,52,62,43,21,44,32,19,50,31,28,25,92,94,95,79,97,80,88,84,89,82,86,81,90,85]
Demographic	Gender, nationality, place of birth, age,	26	[16,27,30,50,68,19,20,51,62,34,43,21,44,49,25,92,93,95,79,84,89,82,86,81,90,85]
Behavioral	Raised hands, visit resources, school satisfaction, discussion, attend class, answer questions	20	[41,29,30,42,20,62,33,21,19,25,91,95,79,80,87,84,89,83,82,90]
Psychological	Personality, motivation, learning strategies, approach to learning, contextual influence.	10	[41,62,34,21,33,93,79,84,83,85]
Family background	Mother & father Education, family income, location of parents	10	[27,29,62,21,44,19,25,92,79,84]
School environment	School size, medium of instruction, lecturer/teacher behavior in class	9	[68,20,33,21,62,19,25,79,83]

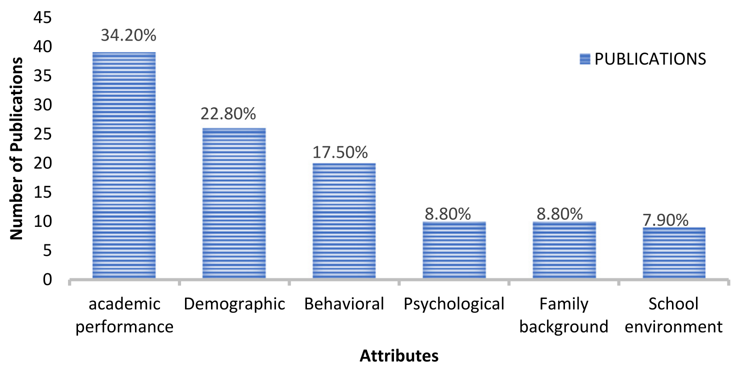

Fig. 4.1. Count of attributes used by authors for performance prediction within the study period.

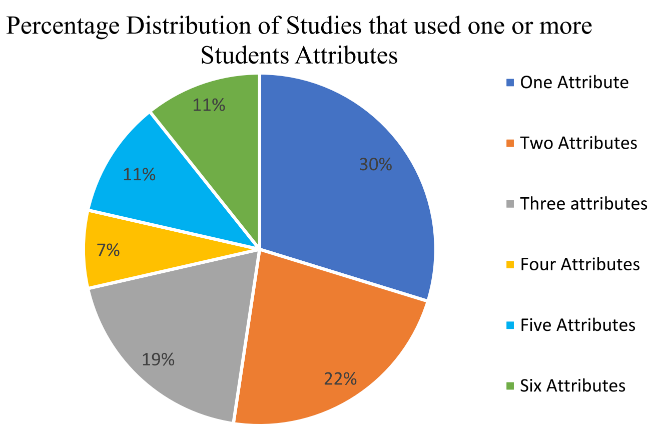

Fig. 4.2. Percentage Distribution of attributes used for student performance within study period.

came in second with 22.80%, followed by behavioral features with 17.50%. The school environment had the fewest occurrences of 7.90%, while the psychological traits and family background each accounted for 8.80% each.

Fig. 4.1 suggests that the majority of researchers believe that indicators of academic accomplishment, such as GPA, are crucial determiners of students' success.

Per indications from Fig. 4.2, using more than one attribute seems to be advantageous for prediction accuracy. 22% of the authors in this review merged two distinct groups of attributes in their investigations, followed by 19% for three attributes and 7% for four attributes. Only 11% of the research fell into the fifth and sixth combinations of attributes, whereas 30% of the studies fell within the category of using one attribute. The importance of using multiple categories of characteristics cannot be overstated, as 70% of the authors used more than one category of characteristics, regardless of whether the publications were for senior high schools or higher education institutions.

Also, according to Fig. 4.1, performance indicators have shown that the best attributes occur most frequently (academic performance = 34.20%). Also, as seen in Table 4.1, a variety of variables have been found to be the best indicators of academic performance. Despite the fact that just two [64,55] studies out of 39 studies in the academic performance domain made this discovery. Majority of the study revealed that students' grades and scores on exams, quizzes, and tests are the factors that have the most influence on their performance. This defies a prior review article in [98] which asserts that the GPA was the most significant factor in predicting student performance. This is true because the main methods for evaluating academic success are grades, and scores from examinations and quizzes as they eventually constitute the GPA. Again, the evaluation of the students' performance cannot solely depend on their grade point average because this review also includes high school and secondary school levels performance, which mostly do not use the GPA system.

Despite having a recurrence rate of 22%, which is relatively high compared to other features, demographic features attributes usage still remains lower than features of academic achievement. Three publications made the observation that, demographic factors such as gender, caste, and parents' socio-economic background have a stronger influence on prediction [30,50,68].

Even though very few studies place a strong emphasis on using school design elements and school environment, this review evaluation was able to locate 9 studies that found that school size [36,47] and the proportion of lecturer attendance [70] were the most important factors in their research.

Table 4.2
Distribution of publications among the three widely used methods for performance prediction per method.

Features	Method per publication	Count	Reference paper
Academic performance	Classification	42	[35,29,58,71,46,23,53,18,22,59,74,73,39,16,57,36, 38,66,47,20,44,21,62,32,52,19,31,30,48,28,25,91, 79,97,87,88,84,83,82,86,81,85]
	Regression	12	[64,58,40,24,54,70,56,59,73,88–90]
	Clustering	6	[55,45,16,57,51,80]
Demographic	Classification	20	[58,71,53,22,59,57,36,66,47,44,21,34,62,19,50,30, 48,25,86,85]
	Regression	9	[37,58,54,70,18,59,49,79,84]
	Clustering	2	[16,57]
Behavioral	Classification	23	[71,1,23,53,18,22,57,36,66,47,34,33,62,19,42,49, 25,92–95,84,86]
	Regression	3	[70,42,49]
	Clustering	1	[57]
Psychological	Classification	13	[29,71,53,22,57,36,66,47,34,62,19,25,85]
	Regression	2	[70,39]
	Clustering	1	[57]
Family socio-economic background	Classification	16	[27,29,71,60,18,22,59,57,36,66,47,44,21,62,83,86]
	Regression	5	[69,70,56,59,19]
	Clustering	1	[57]
School environment	Classification	16	[71,23,53,18,22,57,36,47,20,44,21,62,19,28,25,86]
	Regression	1	[70]
	Clustering	1	[57]

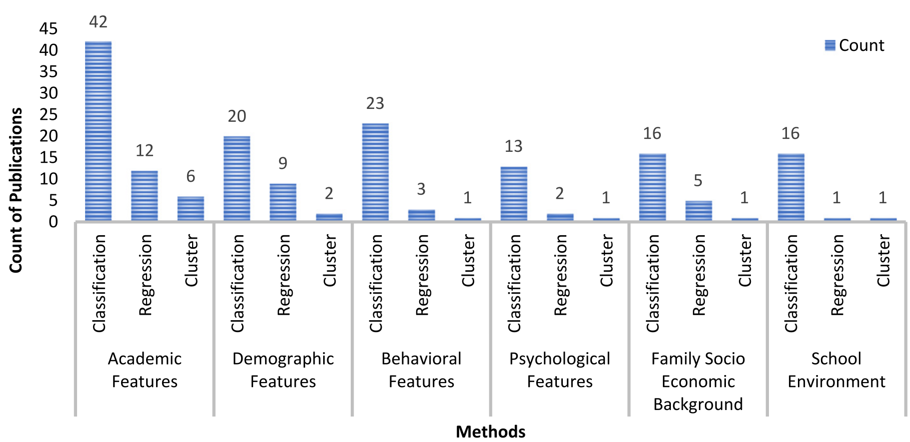

Fig. 4.3. Distribution of Methods by publications on student performance within study period.

Researchers should look beyond traditional characteristics such as academic performance because behavioral features like list of websites visited, visit duration to the library, time spent in watching movies online, and time spent on studying online might help predict students' performance [71].

The features used in this review analysis are shown in Fig. 4.1. The six groups that have been outlined in Table 4.1 were used to split the features. Some features' that were found redundant due to their consolidation with others were renamed and integrated accordingly.

RQ.2: Which methods are frequently employed by researchers when predicting student learning outcomes?

Fig. 4.3 illustrates how various researchers have used algorithms from the clustering, regression, and classification methods, to predict students' success. Whereas regressions are employed in order to discover the connection between dependent and independent variables, classification attempts to properly determine the class labels for every data scenario [99]. Clustering, as opposed to regression and classification problems, is an unsupervised classification approach used to group related items in a dataset. This research found that the classification method had the highest application between the two dominant attributes for performance prediction (42 and 20 counts for academic and demographics respectively). The next most widely used method is regression with (12 and 9 counts for academic and demographics respectively). This indicates that the classification method is the dominantly applied technique for predicting student achievement. This is due to the fact that supervised learning, also known as the classification method, uses labeled data to train algorithms, making it computationally simpler than other approaches. These labeled datasets are also used to train algorithms, which eventually learn to appropriately categorize data or predict outcomes. (See Table 4.2.)

RQ3: Which algorithms or techniques are most effective for predicting student performance?

This literature survey employs a wide range of algorithms. This suggests that there might be other ways to put prediction models into practice. Additionally, a number of models are frequently compared within the same study to determine the model that best fit for a given dataset. Among the algorithms widely employed by academics to estimate student achievement are, Decision Tree (DT), Bayesian Networks (NB), Artificial Neural Network (ANN), Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) [4]. The following are brief descriptions and explanations of the algorithms or methodologies being used in predicting student achievement.

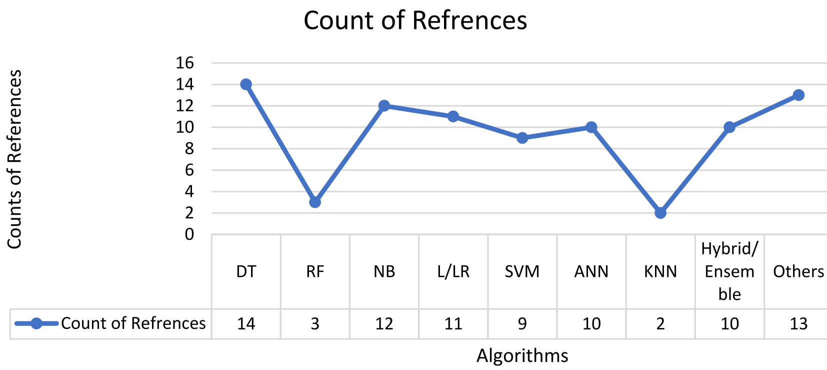

Fig. 4.4. Distribution of algorithms used in the selected papers on students' performance.

Table 4.3
Count of algorithms used in selected publications for performance prediction of students.

Algorithms	References	Count of references
Decision Tree	[20,47,58,73,29,36,22,28,30,52,62,44,79,100]	14
Random Forest	[101,42,57]	3
Naïve Bayes	[5,46,66,57,64,37,31,19,51,32,29,85]	12
Logistic/Linear Regression	[24,69,73,43,40,54,23,50,89,86,90]	11
Support Vector Machine	[43,102,38,1,27,49,95,88,82]	9
Artificial Neural Network	[39,53,77,71,56,48,33,87,81,96]	10
KNN	[55,57],	2
Hybrid/Ensemble	[35,16,74,70,36,59,92–94,97]	10
Other (e.g., ID3, FN, RI)	[69,66,2,59,76,18,25,21,48,91,80,84,83]	13

4.1. The Random Forest (RF) algorithm

As a supervised ensemble ML algorithm, RF is frequently used in classification and regression problems. RF works by first training a decision tree and then creating class output, which represents mode of isolated tree. This analysis discovered that three of the eighty-four papers evaluated used the RF classifier in their datasets for prediction.

4.2. The Support Vector Machine (SVM) algorithm

SVM algorithm is used in education for tracking learner involvement and engagement in courses online. In the majority of applications of machine learning, it has been acknowledged as among the most trustworthy and effective algorithms [99]. In terms of the number and frequency of usage in this study, SVM ranks sixth. SVM was employed in 9 out of the 84 papers to predict student academic results, as indicated in Fig. 4.4.

4.3. The Artificial Neural Network (ANN) algorithm

ANN algorithm was created to simulate the functioning of Biological neural networks. A neural network uses a connectionist method of computing to process information and is composed of a network of artificial neurons that are linked to one another. The connection weights are typically learned by the network using accessible training patterns. The network's performance can be enhanced over time by repetitively updating the network's weights. 10 of the 84 articles that employed the ANN method were used to successfully predict student performance.

4.4. The Logistic and Linear Regression (L/LR) algorithm

In contrast to logistic regression, which forecasts the probabilities of two or more results and permits categorical estimations, linear regression forecasts on a continuous valued result from a linear set of traits. Table 4.3 lists 11 out of 84 articles that employed logistic regression approaches and had successful results.

4.5. The Decision Tree (DT) algorithm

Building the tree and pruning it are the first two steps in the DT classification approach [103]. Branches from the inside nodes of the tree shows the results of a test or condition, while the outward nodes or leaves of the tree indicate class labels [104]. A well-known algorithm for predictive modeling on education-based data is the DT [105]. 14 of the 84 publications that were examined, employed the DT method. The decision tree algorithms were able to outperform all other algorithms when accuracy is considered. The DT is the topmost employed algorithm used among the publications captured in this study.

4.6. The Naïve Bayes (NB) algorithm

The NB algorithm makes learning simple by assuming that variables are autonomous of a specific class while offering a probabilistic interpretation of classification [106]. Though autonomy is an unsuitable assumption in general, the NB classifier frequently outperforms more advanced classifiers in practice. While employing the NB to analyze university students' performance, [107] found that NB algorithm performs well when used for predicting the achievements of basic school pupils in the same task. 12 of the 84 articles under study used the NB algorithms for their predictions.

4.7. The K-Nearest Neighbor (KNN)

The KNN is the least deployed algorithm encountered in the review. The KNN belongs to the Semi-Supervised Machine Learning methods. In this study, only 2 publications were accessed.

4.8. The ensemble/hybrid algorithms

The difficult issue in prediction models is determining which efficient techniques will produce an acceptable prediction accuracy [108]. To obtain the highest level of accuracy, several researchers have proposed the hybrid/ensemble technique, which combines a number of ML algorithms. "Hybrid/ensemble techniques" are defined as the combination of several potential ML algorithms [4]. 10 studies in this work used hybrid algorithm techniques to analyze at-risk students in a course and predict student achievement [35,16,74].

4.9. Other algorithms

Other than the aforementioned algorithms, 13 publications reported good accuracy rates for other algorithms such as ID3, rule induction (RI), and Fuzzy Network (FN) methods.

4.10. Gaps identified in papers reviewed

Population gap

A population gap is a well-known disparity among researchers. There are always neglected populations that have remained understudied or not well represented on an evidential basis or earlier study (e.g., gender, race/ethnicity, age.) [109].

In this SLR study, tasks involving feature engineering, where the kinds of features used might impact how well a classifier performs [24], have received relatively little attention. Students' academic performance, previous schooling, some personal traits of students, and e-learning interaction activity logs have been the four main features employed in the publications under review.

Additionally, the dynamic structure of student achievement is not taken into account in the present literature. The performance of learners is a dynamic process that either increases or decreases with time. Predictor effectiveness on real-time dynamic data needs to be given due attention.

Practical knowledge gap

A gap in practical knowledge occurs when professionalism or practices differ from study findings or are not sufficiently addressed by research [109].

In this SLR, many of the reviewed papers [19] failed to further prescribe a remedy to avert students drop out by using identifiable attributes at their early stages in their programs of study.

Hence there is the need for a comprehensive use of feature engineering on dynamic data of students to track their performance and an early intervention scheme aimed at mapping student features and performance to improve student achievements in schools.

4.11. Significance of the study

The study has been able to identify the following insights in literature pertaining to applying ML on students' data to determine influential factors on their performance.

The SLR demonstrates that, students' prior academic performance is the most used attributes for students' academic performance prediction. Again, the study indicates that, classification method is widely used for classifying problems in academic establishments. Also, the DT had the highest frequency of usage as ML algorithm in students' performance predictions. The study also highlights gaps in both population coverage and practical knowledge in studies relating to ML and students' performance prediction.

4.12. Suggestions for education administrators and researchers

Educational Management Information Systems (EMIS) has a considerable collection of attributes regarded as more relevant by domain experts. However, examining the literature reveals that several attributes are omitted from the past studies. Capturing crucial characteristics such as class attendance, family size, disabilities, teachers' competence, social media network usage, psychological factors, and study methods may aid in appropriately classifying the student. As a result, education institutions and researchers need to scale up the size of their database systems and datasets respectively in order to capture more student related data across all levels of education. This will subsequently close the gap in population coverage realized in past studies.

Again, education administrators and other stakeholders should use the recommended possible set of characteristics of students to implement appropriate strategies (academic intervention schemes) to address students' academic weaknesses as suggested in literature. This can be done by deploying the suggested models by researchers to identify potential traits that will hinder a student's performance there by closing the gap in practical knowledge.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or non-for -profit sectors.

References

[1] B. Owusu-Boadu, I.K. Nti, O. Nyarko-Boateng, J. Aning, V. Boafo, Academic performance modelling with machine learning based on cognitive and non-cognitive features, Appl. Comput. Syst. 26 (2) (2021) 122–131.
[2] A. Siddique, et al., Predicting academic performance using an efficient model based on fusion of classifiers, Appl. Sci. (2021).
[3] B. Albreiki, N. Zaki, H. Alashwal, A systematic literature review of student' performance prediction using machine learning techniques, Educ. Sci. 11 (9) (2021).
[4] H. Nawang, M. Makhtar, W.M.A.F.W. Hamzah, A systematic literature review on student performance predictions, Int. J. Adv. Technol. Eng. Explor. 8 (84) (2021) 1441–1453.
[5] M. Makhtar, H. Nawang, Analysis on students performance using naïve Bayes classifier, in: Analysis on Students Performance using NAÏvE, no. July, 2017, p. 2021.
[6] F. Inusah, Y.M. Missah, U. Najim, F. Twum, Data Mining and Visualisation of Basic Educational Resources for Quality Education, vol. 70, (12) 2022, pp. 296–307.
[7] Y. Baashar, Predicting Student's Performance using Machine Learning Methods : A Systematic Literature Review no. August, 2021.
[8] A.A. Enughwure, M.E. Ogbise, Application of Machine Learning Methods to Predict Student Performance : A Systematic Literature Review, no. May, 2020, pp. 3405–3415.
[9] S.N. Singh, Educational Data Mining and Its Role in Determining Factors Affecting Students Academic Performance : A Systematic Review, no. August, 2016.
[10] L. Sandra, F. Lumbangaol, T. Matsuo, Machine Learning Algorithm to Predict Student's Performance: A Systematic Literature Review, vol. 10, (4) 2021, pp. 1919–1927.
[11] B. Sekeroglu, R. Abiyev, A. Ilhan, M. Arslan, J.B. Idoko, Systematic literature review on machine learning and student performance prediction : Critical gaps and possible remedies, Appl. Sci. (2021).
[12] F. Ofori, E. Maina, R. Gitonga, Using machine learning algorithms to predict students' performance and improve learning outcome: A literature based review francis ofori, dr. Elizaphan maina and dr. Rhoda gitonga, using machine learning algorithms to predict students, J. Inf. Technol. (ISSN: 2617-3573) 4 (1) (2020) 33–55.
[13] P. Balaji, S. Alelyani, A. Qahmash, Contributions of machine learning models towards student academic performance prediction : A systematic review, Appl. Sci. (2021).
[14] B. Kitchenham others, Systematic literature reviews in software engineering – A tertiary study, Inf. Softw. Technol. 52 (8) (2010) 792–805.
[15] C. Okoli, A Guide to Conducting a Standalone Systematic, vol. 37, 2015.
[16] M. Pandey, S. Taruna, Towards the integration of multiple classifier pertaining to the student's performance prediction, Perspect. Sci. 8 (2016) 364–366.
[17] P.G. Sameer, S.R. Barahate, Educational Data Mining – A New Approach to the Education Systems, 2016, pp. 18–20.

[18] P. Chair others, Organizing Committee General and Financial Chair Organizing Secretary Technical Programme Committee Advisory Committee Predicting and Analyzing Students' Performance: An Educational Data Mining Approach no. September, 2016.

[19] C. Anuradha, T. Velmurugan, Feature selection techniques to analyse student acadamic performance using naïve bayes classifier, in: 3rd Int. Conf. Small Mediu. Bus, 2016, pp. 345–350.

[20] R. Sumitha, E.S. Vinothkumar, P.G. Scholar, Prediction of students outcome using data mining techniques, Int. J. Sci. Eng. Appl. Sci. 2 (6) (2016) 132–139.

[21] C. Anuradha, T. Velmurugan, Fast Boost Decision Tree Algorithm: A Novel Classifier for the Assessment of Student Performance in Educational Data, vol. 31, 2016, pp. 254–0223.

[22] B. Mehboob, R.M. Liaqat, N.A. Saqib, Predicting student performance and risk analysis by using data mining approach, Int. J. Comput. Sci. Inf. Secur. 14 (7) (2016) 69–76.

[23] F. Keshtkar, J. Cowart, A. Crutcher, Predicting risk of failure in online learning platforms using machine learning algorithms for modeling students' academic performance, in: 33rd Int. Conf. Mach. Learn, vol. 48, 2016.

[24] Neelam Peters, Aakanksha S. Choubey, A survey on data classification and machine learning for forecasting of student performance, Int. J. Eng. Sci. Res. Technol. 5 (12) (2016) 934–940, http://dx.doi.org/10.5281/zenodo.222225.

[25] M. Peker, Predicting and Analyzing Students' Performance: An Educational Data Mining Approach, 2016, pp. 36–40.

[26] N. Peters, A.S. Choubey, A survey on data classification and machine learning for forecasting of student performance, Int. J. Eng. Scieneces Res. Technol. 5 (12) (2016) 934–940.

[27] A. Daud, Predicting Student Performance using Advanced Learning Analytics, 2017, pp. 415–421.

[28] R. Asif, A. Merceron, S.A. Ali, N.G. Haider, Analyzing undergraduate students' performance using educational data mining, Comput. Educ. 113 (2017) 177–194.

[29] B. Gen, S. Salahuddin, H.K. Talukder, Original Article Influence of Socio-Demographic Characteristics on Academic Performance of Medical Students, vol. 8, (2) 2017, pp. 18–23.

[30] M. Pojon, T1_Using machine learning to predict student performance, no. June, Univ. Tamp. 1–28.

[31] M. Makhtar, H. Nawang, Analysis on Students Performance using NaÏve Bayes Classifier analysis on students performance using NAÏvE no. August, 2017.

[32] T. Case, Student performance prediction model using machine learning approach, no. February, Int. J. Adv. Res. Stud. Perform. Predict. Model Mach. Learn. Approach (2017) The Case of Wolkite University.

[33] L. Gerritsen, Predicting Student Performance with Neural Networks no. May, 2017, pp. 1–30.

[34] D. Schreiber, The Impact of Demographic Influences on Academic Performance and Student Satisfaction with Learning as Related to Self-Esteem, Self- Efficacy and Cultural Adaptability Within the Context of the Military, vol. 16, (4) 2017, pp. 67–90.

[35] O.W. Adejo, T. Connolly, Predicting Student Academic Performance using Multi-Model Heterogeneous Ensemble Approach no. December, 2017.

[36] S.K. S. Agrawal, A. K, Using data mining classifier for predicting student's performance in UG level, Int. J. Comput. Appl. 172 (8) (2017) 39–44.

[37] A.K. Hamoud, A.S. Hashim, Students ' Success Prediction Based on Bayes Algorithms, vol. 178, (7) 2017, pp. 6–12.

[38] S.A. Oloruntoba, Student academic performance prediction using support vector machine, IJESRT Int. J. Eng. Sci. Res. Technol. 6 (12) (2017) 588–598.

[39] R.R. Halde, A. Deshpande, A. Mahajan, Psychology assisted prediction of academic performance using machine learning, in: 2016 IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol., RTEICT 2016 - Proc. (2017), 2016, pp. 431–435.

[40] O.D. Oyerinde, Predicting Students ' Academic Performances – A Learning Analytics Approach using Multiple Linear Regression, vol. 157, (4) 2017, pp. 37–44.

[41] C. Simon, Y. Bugusa, Survey on data mining approach for analysis and prediction of student performance, Int. J. Eng. Technol. 7 (4.5) (2018) 467.

[42] M.M.Z. Eddin, N.A. Khodeir, H.A. Elnemr, A comparative study of educational data mining techniques for skill-based predicting student performance, Int. J. Comput. Sci. Inf. Secur. 16 (3) (2018) 56–62.

[43] S. Vector, R.A. Case, E.Y. Obsie, S.A. Adem, Prediction of Student Academic Performance using Neural Network, Linear Prediction of Student Academic Performance using Neural Network, Linear Regression and Support Vector Regression : A Case Study, 2018, p. 2019.

[44] R. Kaur, Use of data mining to predict slow learners in senior secondary schools, IJESRT Int. J. Eng. Sci. Res. Technol. 186 (4) (2018) 186–191.

[45] P. Nor, E. Nohuddin, Z. Zainol, A. Nordin, Monitoring students performance using self organizing map trend clustering, ZULFAQAR Int. J. Def. Sci. Eng. Technol. 1 (1) (2018) 50–56.

[46] M. Studies, T. Nadu, Prediction Accuracy on Academic Performance of Students using Different Data Mining Algorithms with Influencing Factors, vol. 7, (5) 2018.

[47] R. Patil, S. Tamane, A comparative analysis on the evaluation of classification algorithms in the prediction of diabetes, Int. J. Electr. Comput. Eng. 8 (5) (2018) 3966–3975.

[48] F. Junshuai, Predicting Students' Academic Performance with Decision and Neural Network, a$\gamma\alpha\eta$, vol. 8, (5) 2019, p. 55.

[49] B. Sekeroglu, K. Dimililer, K. Tuncal, Student performance prediction and classification using machine learning algorithms, PervasiveHealth Pervasive Comput. Technol. Healthc. Part F1481 (2019) 7–11.

[50] A.I. Adekitan, E. Noma-Osaghae, Data mining approach to predicting the performance of first year student in a university using the admission requirements, Educ. Inf. Technol. 24 (2) (2019) 1527–1543.

[51] W.F.W. Yaacob, S.A.M. Nasir, W.F.W. Yaacob, N.M. Sobri, Supervised data mining approach for predicting student performance, Indones. J. Electr. Eng. Comput. Sci. 16 (3) (2019) 1584–1592.

[52] D. Buenaño-fern, D. Gil, Application of Machine Learning in Predicting Performance for Computer Engineering Students : A Case Study, 2019, pp. 1–18.

[53] V.K. Pal, V.K.K. Bhatt, Performance prediction for post graduate students using artificial neural network, Int. J. Innov. Technol. Explor. Eng. 8 (7) (2019) 446–454.

[54] R.R. Rajalaxmi, P. Natesan, N. Krishnamoorthy, S. Ponni, Regression model for predicting engineering students academic performance, Int. J. Recent Technol. Eng. 7 (6) (2019) 71–75.

[55] L. Mahmoud, A. Zohair, Prediction of Student's Performance By Modelling Small Dataset Size, 2019.

[56] E.T. Lau, L. Sun, Q. Yang, Modelling, prediction and classification of student academic performance using artificial neural networks, SN Appl. Sci. 1 (9) (2019) 1–10.

[57] Y. Widyaningsih, N. Fitriani, D. Sarwinda, A semi-supervised learning approach for predicting student's performance: First-year, in: 2019 12th Int. Conf. Inf. Commun. Technol. Syst., 2019, pp. 291–295.

[58] O.G. Funmilayo, O. State, O. State, Student's Performance Prediction Using Multiple Linear Regression and Decision Tree, vol. 8, (7) 2019, pp. 256–268.

[59] F. Aman, A. Rauf, R. Ali, F. Iqbal, A.M. Khattak, A predictive model for predicting students academic performance, in: 10th Int. Conf. Information, Intell. Syst. Appl. IISA, vol. 2019, 2020, p. 2019.

[60] J. David, G. Anastasija, Predicting Academic Performance Based on Students' Family Environment: Evidence for Colombia Using Classification Trees, vol. 11, (3) 2019, pp. 299–311.

[61] C. Romero, S. Ventura, Educational data mining and learning analytics: An updated survey, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 10 (3) (2020) 1–21.

[62] D.R. Kawade, K.S. Oza, P.G. Naik, Student performance classification: A data mining approach, JIMS8I Int. J. Inf. Commun. Comput. Technol. 8 (2) (2020) 462–466.

[63] E. Alyahyan, D. Düştegör, Predicting academic success in higher education: Literature review and best practices, Int. J. Educ. Technol. High. Educ. 17 (1) (2020).

[64] E. Edu, C.C. Jr., J.K.D. Treceñe, Application of Machine Learning Algorithms in Predicting the Performance of Students in Mathematics in the Modern World, vol. 1, (1) 2020, pp. 49–57.

[65] R. Tabassum, N. Akhter, Effect of Demographic Factors on Academic Performance of University Students, vol. 14, (1) 2020, pp. 64–80.

[66] D.T. Ha, C.N. Giap, P.T.T. Loan, T.L.H. Huong, An empirical study for student academic performance prediction using machine learning techniques, Int. J. Comput. Sci. Inf. Secur. 18 (3) (2020) 21–28.

[67] A. Bilal Zorić, Benefits of educational data mining, J. Int. Bus. Res. Mark. 6 (1) (2020) 12–16.

[68] G.A. El Refae, A. Kaba, S. Eletter, The Impact of Demographic Characteristics on Academic Performance: Face-to-Face Learning Versus Distance Learning Implemented to Prevent the Spread of COVID, vol. 22, (1) 2021.

[69] R. Gupta, C. Gueneau, Feature Correlation with Student Education Performance, vol. 10, (2) 2021, pp. 1–13.

[70] M.N. Yakubu, A.M. Abubakar, Applying machine learning approach to predict students' performance in higher educational institutions, no. June, Kybernetes (2021).

[71] B.K. Yousafzai, S.A. Khan, T. Rahman, I. Khan, I. Ullah, A.U. Rehman, Student-Performulator: Student Academic Performance using Hybrid Deep Neural Network, 2021, pp. 1–21.

[72] J. López-Zambrano, J.A.L. Torralbo, C. Romero, Early prediction of student learning performance through data mining: A systematic review, Psicothema 33 (3) (2021) 456–465.

[73] S. Hussain, M.Q. Khan, Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning, no. Ml, Ann. Data Sci. (2021).

[74] F. Saleem, Z. Ullah, B. Fakieh, F. Kateb, Intelligent decision support system for predicting student's e-learning performance using ensemble machine learning, Mathematics 9 (17) (2021) 739–749.

[75] M.A. Saleh, S. Palaniappan, N.A.A. Abdalla, Education is an overview of data mining and the ability to predict the performance of students, Edukasi 15 (1) (2021) 19–28.

[76] M.N. Yakubu, Applying Machine Learning Approach to Predict Students' Performance in Higher Educational Institutions no. June, 2021.

[77] S. Kisnanto, R.R.R. Hadiani, C. Ikhsan, Students ' Performance Analyses using Machine Learning Algorithms in WEKA Students ' Performance Analyses using Machine Learning Algorithms in WEKA, 2021.

[78] P. Dabhade, R. Agarwal, K.P. Alameen, A.T. Fathima, R. Sridharan, G. Gopakumar, Educational data mining for predicting students' academic performance using machine learning algorithms, Mater. Today: Proc. 47 (xxxx) (2021) 5260–5267.

[79] D.M. Ahmed, A.M. Abdulazeez, D.Q. Zeebaree, F.Y.H. Ahmed, Predicting university's students performance based on machine learning techniques, in: 2021 IEEE Int. Conf. Autom. Control Intell. Syst, I2CACIS 2021 - Proc., no. August, 2021, pp. 276–281.

[80] Z. Mingyu, W. Sutong, W. Yanzhang, W. Dujuan, An interpretable prediction method for university student academic crisis warning, Complex Intell. Syst. 8 (1) (2022) 323–336.

[81] S. Trivedi, Improving students' retention using machine learning: Impacts and implications, no. August, Sci. Prepr. (2022) 0–2.

[82] F. Ouatik, M. Erritali, F. Ouatik, M. Jourhmane, Predicting student success using big data and machine learning algorithms, Int. J. Emerg. Technol. Learn. 17 (12) (2022) 236–251.

[83] E. Melo, I. Silva, D.G. Costa, C.M.D. Viegas, T.M. Barros, On the use of explainable artificial intelligence to evaluate school dropout, Educ. Sci. 12 (12) (2022).

[84] M. Tadese, A. Yeshaneh, G.B. Mulu, Determinants of good academic performance among university students in Ethiopia: A cross-sectional study, BMC Med. Educ. 22 (1) (2022) 1–9.

[85] S. Verma, R.K. Yadav, K. Kholiya, Prediction of academic performance of engineering students by using data mining techniques, Int. J. Inf. Educ. Technol. 12 (11) (2022) 1164–1171.

[86] M.N. Yakubu, A.M. Abubakar, Applying machine learning approach to predict students' performance in higher educational institutions, Kybernetes 51 (2) (2022) 916–934.

[87] R.K. Veluri others, Learning analytics using deep learning techniques for efficiently managing educational institutes, Mater. Today: Proc. 51 (2022) (2022) 2317–2320.

[88] M. Yağcı, Educational data mining: Prediction of students' academic performance using machine learning algorithms, Smart Learn. Environ. 9 (1) (2022).

[89] A.S. Hashim, W.A. Awadh, A.K. Hamoud, Student performance prediction model based on supervised machine learning algorithms, IOP Conf. Ser. Mater. Sci. Eng. 928 (3) (2020).

[90] J. Neroni, C. Meijs, P.A. Kirschner, K.M. Xu, R.H.M. de Groot, Academic self-efficacy, self-esteem, and grit in higher online education: Consistency of interests predicts academic success, Soc. Psychol. Educ. 25 (4) (2022) 951–975.

[91] V. Swamy, B. Radmehr, N. Krco, M. Marras, T. Käser, Evaluating the Explainers: Black-Box Explainable Machine Learning for Student Success Prediction in MOOCs, 2022.

[92] M. Arashpour others, Predicting individual learning performance using machine-learning hybridized with the teaching-learning-based optimization, Comput. Appl. Eng. Educ. 31 (1) (2023) 83–99.

[93] K. Okoye, A. Arrona-Palacios, C. Camacho-Zuñiga, J.A.G. Achem, J. Escamilla, S. Hosseini, Towards Teaching Analytics: A Contextual Model for Analysis of Students' Evaluation of Teaching Through Text Mining and Machine Learning Classification, vol. 27, (3) 2022.

[94] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, P.C. Nshimyumukiza, Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization,, Comput. Educ. Artif. Intell. 3 (2021) (2022) 100066.

[95] K. Sixhaxa, A. Jadhav, R. Ajoodha, Predicting students performance in exams using machine learning techniques, in: Proc. Conflu. 2022-12th Int. Conf. Cloud Comput. Data Sci. Eng, 2022, pp. 635–640.

[96] A. Alhothali, M. Albsisi, H. Assalahi, T. Aldosemani, Predicting student outcomes in online courses using machine learning techniques: A review, Sustain. 14 (10) (2022) 1–23.

[97] Z. Kanetaki, C. Stergiou, G. Bekas, C. Troussas, C. Sgouropoulou, A hybrid machine learning model for grade prediction in online engineering education, Int. J. Eng. Pedagog. 12 (3) (2022) 4–23.

[98] A.M. Shahiri, W. Husain, N.A. Rashid, A review on predicting student's performance using data mining techniques, Procedia Comput. Sci. 72 (2016) (2015) 414–422.

[99] H. Hu, G. Zhang, W. Gao, Big data analytics for MOOC video watching behavior based on spark, Neural Comput. Appl. (2019) 0123456789.

[100] M.A. Verma, An Insight Into Self-Efficacy and Its Impact on Students ' Achievement- A Review, vol. 6, (4) 2022, pp. 2746–2752.

[101] J. Kaunang, Students' academic performance prediction using data mining, in: Third Int. Conf. Informatics Comput, 2018, pp. 1–5.

[102] I. Journal, Student Academic Performance Prediction Using Support Vector Machine.

[103] A.K. Hamoud, A.S. Hashim, W.A. Awadh, Predicting student performance in higher education institutions using decision tree analysis, Int. J. Interact. Multimed. Artif. Intell. 5 (2) (2018) 26.

[104] M.S. Zulfiker, N. Kabir, A.A. Biswas, P. Chakraborty, M.M. Rahman, Predicting students' performance of the private universities of Bangladesh using machine learning approaches, Int. J. Adv. Comput. Sci. Appl. 11 (3) (2020) 672–679.

[105] S. Sivakumar, S. Venkataraman, R. Selvaraj, Predictive modeling of student dropout indicators in educational data mining using improved decision tree, Indian J. Sci. Technol. 9 (4) (2016) 1–5.

[106] E.P.F. Lee, et al., An ab initio study of rbo, CsO and FrO (x2∑+; A2∏) and their cations (x3∑-; A3∏), Phys. Chem. Chem. Phys. 3 (22) (2001) 4863–4869.

[107] D. Gašević, V. Kovanović, S. Joksimović, Piecing the learning analytics puzzle: A consolidated model of a field of research and practice, Learn. Res. Pract. 3 (1) (2017) 63–78.

[108] P. Sokkhey, T. Okazaki, Hybrid machine learning algorithms for predicting academic performance, Int. J. Adv. Comput. Sci. Appl. 11 (1) (2020) 32–41.

[109] D.A. Miles, A Taxonomy of Research Gaps: Identifying and Defining the Seven Research Gaps no. August, 2017.