Journal of Information Technology Education: Research

An Official Publication of the Informing Science Institute InformingScience.org

JITEResearch.org

Volume 24, 2015

A Systematic Review of Machine Learning Techniques for Predicting Student Engagement in Higher Education Online Learning

Ke Ting Chong *	Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia	keting.phd@gmail.com keting-1997@graduate.utm.my
Noraini Ibrahim	Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia	noraini_ib@utm.my
Sharin Hazlin Huspi	Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia	sharin@utm.my
Wan Mohd Nasir Wan Kadir	Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia	wnasir@utm.my
Mohd Adham Isa	Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia	mohdadham@utm.my

* Corresponding author

Abstract

Aim/Purpose	The purpose of this study is to review and categorize current trends in student engagement and performance prediction using machine learning techniques during online learning in higher education. The goal is to gain a better understanding of student engagement prediction research that is important for current educational planning and development. However, implementing machine learning approaches in student engagement studies is still very limited.
Background	The rise of online learning during and after COVID-19 has created new difficulties for students' engagement and academic achievements. Lecturers' manual monitoring and supporting of students are inadequate online, leading to

Accepting Editor Fariza Khalid │ Received: September 10, 2024 │ Revised: November 27, December 23, 2024 │ Accepted: February 7, 2025.

Cite as: Chong, K. T., Ibrahim, N., Huspi, S. H., Wan Kadir, W. H. N., & Isa, M. A. (2025). A systematic review of machine learning techniques for predicting student engagement in higher education online learning. *Journal of Information Technology Education: Research, 24,* Article 5. https://doi.org/10.28945/5456

(CC BY-NC 4.0) This article is licensed to you under a Creative Commons Attribution-NonCommercial 4.0 International License. When you copy and redistribute this paper in full or in part, you need to provide proper attribution to it to ensure that others can later locate this work (and to ensure that others do not accuse you of plagiarism). You may (and we encourage you to) adapt, remix, transform, and build upon the material for any non-commercial purposes. This license does not permit you to use this material for commercial purposes.

disengagement and performance challenges that may be very difficult to notice. Machine learning has great potential to provide an accurate prognosis of students' engagement and outcomes to make early interventions possible. Nevertheless, the current knowledge deficit is in the systematic presentation of trends and insights concerning the utilization of these approaches in higher education online learning, especially with a focus on student engagement research. This research fills a crucial void by explaining and analyzing current trends in machine learning-based prediction models to enhance the quality and efficiency of an online learning system.

Methodology This research examines the existing literature on the application of machine learning, which allows computers to learn from data and improve their performance for early identification of student engagement and academic performance in higher education during online learning. The PICOC protocol was implemented to guide the search process and define the relevant keywords aligned with the research questions. Based on the PRISMA framework, a structured approach is adopted to identify and select studies to screen and extract the relevant papers from the database. Meta-analysis was adopted in data analysis whereby studies are combined and evaluated to provide insights into machine learning techniques' effectiveness in student engagement and academic performance research.

Contribution This paper aims to present the current trends in predicting student engagement and academic achievement by applying machine learning approaches with a focus on their relevance in the context of online learning. It defines challenges that emerge with an interpretation of the extent of student engagement, which include the absence of consensus on levels of student engagement that hampers the use of explainable artificial intelligence – approaches that make training of machine learning models more logical, understandable and easily interpretable by lecturers. The finding points to the fact that through the prediction models, lecturers are enabled to recognize disengaged students early and foster their needs towards learning, providing direction toward more customized and effective online learning.

Findings A total of 96 primary studies have been identified and included in this systematic review. It is important to highlight the relevance of classification machine learning methods that are implemented in 88.60% of papers, while clustering methods are only employed in 15.19% of studies. Furthermore, the review shows that most research focuses on student performance prediction (82.28%) compared to student engagement level prediction (12.66%). Besides, student engagement datasets are used in 92.14% of studies, emphasizing student engagement's popularity in educational prediction research. Moreover, classification machine learning methods are more prevalent in educational prediction research. In contrast, classification methods for student engagement research are still limited due to challenges in constructing consistent engagement levels.

Recommendations for Practitioners Lecturers need to occasionally assess student engagement levels during online learning to identify students who are left out and take immediate planning and action to encourage the student to engage during online learning. The syllabus designer should observe the students' engagement level during online learning to plan the course content that can attract and engage the students. Students' engagement during online learning can ensure their academic success and prevent them from dropping out.

Recommendations for Researchers	Researchers should focus on defining the consensus on differentiating student engagement levels and implementing more explainable AI to enhance the interpretability and transparency of student engagement level predictive models. Researchers should enhance educational predictive models' explainability, transparency, and accuracy by addressing issues brought about by feature selection, resampling techniques, and hyperparameter tuning.
Impact on Society	The study highlights the growing importance of understanding student engagement through digital footprints, which can support personalized learning experiences and provide better educational outcomes. The efficient predictive models on student engagement can improve the effectiveness of higher education systems, benefiting students and institutions.
Future Research	The challenges of current computational methods need to be overcome, including the need for more consistent approaches in differentiating engagement levels and enhancing the explainability and accuracy of educational predictive models through better feature selection, resampling techniques, and hyperparameter tuning.
Keywords	machine learning, prediction, student engagement, student performance, systematic literature review

INTRODUCTION

Technological enhancement over the years has influenced different sectors of human life and, more specifically, education during the COVID-19 pandemic (Ch'ng, 2024). Educational Data Mining (EDM) is a relatively young area that concerns machine learning (ML) adaptation that helps to find previously unknown patterns and relations in educational data due to the increase of data availability through Learning Management Systems (LMS) (Demong et al., 2023; Feng et al., 2022; S. N. Ismail et al., 2021). Online resources as a part of the learning process are increasing, and it is getting more challenging to find unique data in the pools (Khanal et al., 2020). In today's fast-paced digital technology world, online learning is becoming a crucial form of learning globally (Hsueh et al., 2022). During the COVID-19 outbreak, most higher education institutions (HEIs) were forced to shift from face-to-face to distance learning. However, the implementation of distance learning has encountered low student engagement (SE) challenges (Demong et al., 2023). Therefore, it is necessary to study both SE and behavior to improve the quality of online learning (OL) implementation, which is cost-effective and adaptable but lacks direct student-student and student-instructor interaction (Benabbes et al., 2023; Uliyan et al., 2021).

Explainable Artificial Intelligence (XAI) is particularly valuable in educational research since it offers explainability as well as accuracy in prediction models. Education stakeholders seek high-accuracy prediction and desire insight into the rationale behind predictions (Alamri & Alharbi, 2021). XAI can be described as AI systems that explain their decisions to users in an easy-to-understand and natural language. This paradigm is needed in all domains where AI-generated output needs to be evaluated for transparency and explainability. XAI increases transparency by outlining the causes of AI prediction-making decisions, hence enabling decision-making to adhere to ethical guidelines (Rane & Paramesha, 2024; Yadav, 2024). XAI has made many improvements in transparency and explainability. However, the proper balance between interpretability and model performance is still a crucial topic that needs further research and improvement in the future (Agrawal & Sharma, 2024).

Several studies mention that low student engagement has become a crucial challenge during online learning, especially during the COVID-19 pandemic (Hollister et al., 2022; D. Yang et al., 2023). For example, a study illustrates that students prefer recorded lectures due to their flexibility compared to synchronous learning (D. Yang et al., 2023). However, the recorded lecturer sometimes might minimize students' active participation. In addition, up to 72% of students frequently experience social loneliness and disconnection with peers and lecturers, which also reduces the students' engagement (Hollister et al., 2022). The low SE issues are caused by the lack of social interactions and the increased number of distractions during the implementation of OL. Furthermore, Hollister et al. (2022) highlighted that during the COVID-19 pandemic, student attendance in live sessions reduced significantly, with students being absent from lectures and using only recording. This was more apparent in institutions that implemented the policy of giving lower grades during the time of the pandemic.

Even though OL provides a more flexible learning environment, it is difficult to ensure that students participate at the same level as regular education (Turan & Karabey, 2023). The lecturers' digital competencies are essential in constructing an active online learning environment to ensure students' engagement and decrease the students' sense of isolation (ElSayary et al., 2022). Moreover, students lacking self-learning habits cannot be adequately supported and cannot effectively develop skills and attitudes in practice-based courses (Palanci et al., 2024). This is because self-learning activities are the most crucial for implementing distance learning (Hsueh et al., 2022). In OL, the lack of student engagement was indicated in the research as an antecedent to failure and dropout (Palanci et al., 2024). According to Palanci et al. (2024), most research is concerned with enhancing the learning process, student engagement, and effective teaching and learning (T&L). Meanwhile, artificial intelligence has introduced different methods of enhancing T&L in OL, particularly in academic performance, online participation, and engagement (Ouyang et al., 2022). The lack of personalized educational activities tailored to fit individual student requirements further increases the challenges for effective learning (Demong et al., 2023). SE has emerged as a phenomenon that presents a huge risk to HEIs in terms of disruption of student retention, satisfaction, motivation, and success in OL (Benabbes et al., 2023; Uliyan et al., 2021).

Tai et al. (2020) conducted a systematic literature review (SLR) in which SE was defined as a multidimensional construct that reflects behavioral, cognitive, emotional, and social aspects. S. N. Ismail et al. (2021) suggested the integration of ML in LMS to improve SE assessment, but a critical evaluation of SE levels by applying ML is lacking. In contrast, when the SLR was conducted for student performance (SP) prediction, it was elicited that supervised ML is a widely used method for this purpose (Namoun & Alshanqiti, 2020). Albreiki et al. (2021) noted that there is an absence of sufficient literature focusing on the task of feature engineering, which can be a promising way to improve the performance of the ML models.

Therefore, this research aims to establish gaps and trends in predicting SE levels using SLR. Considering that there is relatively weak empirical evidence on SE prediction in contrast to SP, this paper analyses the trend in both during OL for HEIs. This paper explains what SE is and the present trend of ML approaches that have been adopted to estimate SE and SP, namely classification, clustering, and regression. Furthermore, the present research considers the possible methods that could be used to optimize the performance of the ML model as well as the measures used in assessing the performance of the ML algorithms.

LITERATURE REVIEW

The dataset that is often used for predicting student success and SP is learning and activity engagement (Namoun & Alshanqiti, 2020). Subramainan and Mahmoud (2020) and Tai et al. (2020) describe SE as a construct that is characterized by behavioral, cognitive, emotional, and social elements. Behavioral, emotional, and cognitive dimensions are the most dominant and imply poor engagement and commitment to study (Subramainan & Mahmoud, 2020). In contrast, S. N. Ismail et al. (2021) divide student LMS interactions into student-student, student-lecturer, and student system. Each of these components can be determined by the times students log into LMS, and the appropriate files are downloaded. The data collected proposed several factors defining engagement, including LMSs and the learning design, students' apathy to engage late, duration of tasks, motivation, engagement of educators, and monitoring. The level of self-motivation can be deduced by the level of cognitive, emotional, and behavioral investment of students. Besides, this research envisioned the application of ML strategies in LMS to improve assessment of SE level (S. N. Ismail et al., 2021). Subramainan and Mahmoud (2020) highlighted that there is a need to investigate the emotional part in future research.

Student demographics, academic records, and logs of e-learning interaction sessions are the most employed features for the prediction of SP (Abu Saa et al., 2019; Albreiki et al., 2021; Baashar et al., 2021). Internal assessment, communication, and psychological measures are also the attributes utilized for the prediction of SP. Among all these predictors, academic records are the most widely used to predict SP (Baashar et al., 2021). The implementation of educational psychology theory, which includes self-determination theory (SDT) and motivation-based models, can offer a more refined view of student engagement. The key concepts of engagement for SDT are intrinsic motivation and people's interests in autonomy, competence, and relatedness (Ersozlu et al., 2024; Pooja, 2024). Consequently, by incorporating these psychological constructs into machine learning algorithms, researchers have a chance to develop improved and accurate models for evaluating students' motivational usage of OL platform. Furthermore, the present literature suggests that enhancing knowledge of students' self-regulated learning processes and attitudes can contribute to better learning results (Wang et al., 2021). Hence, future work should focus on the integration of these two philosophies, machine learning with educational psychology, to enhance the formation of holistic predictive models that can enhance learning.

The majority of the SP predictions are conducted by implementing supervised ML. In contrast, there is very limited research implementing unsupervised ML for prediction (Namoun & Alshanqiti, 2020). Among various types of machine learning approaches, conventional machine learning algorithms, such as Support Vector Machine (SVM), Decision Trees (DT), Naïve Bayes (NB), k-Nearest Neighbour (KNN), NN, and ANN, are often employed for the prediction of SP (Abu Saa et al., 2019; Albreiki et al., 2021; Baashar et al., 2021). On the other hand, deep reinforcement learning is implemented to automate personalized recommendations and track student engagement. It is also implemented to categorize students into risk groups based on their performance prediction (Bagunaid et al., 2024). In the meantime, the sequential reinforcement learning engagement models integrate engagement detection with sequential predictions, utilizing LSTM networks to learn from students' engagement patterns and demographic features (Song et al., 2021).

Meanwhile, XAI has important applications in behavioral prediction because it provides a clearer view of the inner workings of artificial intelligence models, especially where user behavior and psychological patterns are concerned. XAI techniques have been used as a successful model for analyzing the issue of adolescents' problematic internet use. Applying the feature selection and importance methods, researchers can successfully explain the predictive models to understand the effects of features on user behavior (Stanimirovic et al., 2024). Furthermore, XAI facilitates the extraction of psychological traits from digital footprints, such as spending data. Techniques like global and local rule extraction help clarify how specific behavior correlates to personality traits, thus allowing for model validation and improvement (Ramon et al., 2021). Besides, the present research lacks emphasis on feature engineering to enhance the performance of the prediction model. Therefore, feature engineering and dealing with class imbalance is the novel future step for upcoming research (Albreiki et al., 2021).

Among all SLRs related to SE, only S. N. Ismail et al. (2021) pay limited attention to ML for prediction, whereas Subramainan and Mahmoud (2020) and Tai et al. (2020) merely discuss the concept of SE. Articles covering SP focus on the utilization of ML algorithms for predicting student outcomes. However, most studies (e.g., Albreiki et al., 2021; Baashar et al., 2021; Namoun & Alshanqiti, 2020) focus mainly on the classification approach, whereas Abu Saa et al. (2019) include clustering ML within their SLR. Recall from the work of Albreiki et al. (2021) that very little feature engineering has been done for SP prediction, where the problem of class imbalance is a significant issue. There is a dearth of studies that deal with feature engineering and class balance in the literature that informs this paper. Hence, this SLR will particularly aim to identify ML to determine SE and SP only. This research will also explore one of the most common practices of feature engineering and class balancing in the context of SE and SP prediction.

RESEARCH METHODOLOGY

This SLR was undertaken to examine studies that are related to SE and SP based on a specific research goal and a structured method. The structured method and process of SLR implemented in this SLR was inspired by Kitchenham (2004) and Khatibsyarbini et al. (2018), which include mainly six different phases: identify research questions, select literature repositories, identify the search string, select the related studies, and synthesize and extract the data.

STEP 1: IDENTIFY RESEARCH QUESTIONS

The aim of this SLR is to comprehend and review recent experimental evidence that reflects the prediction of SE and SP for further investigation, and the final goal of this review is to improve the ability of present studies. To achieve the goal of this study, five research questions are

1. What is student engagement?
2. What is the popular data used to predict student engagement or performance?
3. What are the popular machine learning methods used to predict student engagement or student performance?
4. How to improve the performance of the prediction models?
5. How to measure the prediction performance of a machine learning algorithm?

STEP 2: SELECT LITERATURE REPOSITORIES

There are six online repositories used in this SLR to search for studies. The repositories are selected based on their variation of article resources and their credibility levels. The repositories that were used in this SLR are listed and described in Table 1.

Table 1. Online repositories and reason for selection

Online repository	Reason for selection
ACM	The world's largest educational and scientific computing society advances computing resources.
IEEEXplore	Web access to over 4 million full-text documents from highly cited publications in electrical engineering, computer sciences, and other fields.
ScienceDirect	Leading source of over 250,000 open-access articles and journals in physical science and engineering, social sciences, life sciences, and health sciences
Scopus	Repository of peer-reviewed journals in art and humanities, social sciences, scientific, technical, and medical
Web of Science (WoS)	Repository offering comprehensive citations from journals across the academic and scientific fields, including sciences, social sciences, and technologies.
PLoS ONE	Repository providing peer-reviewed open-access scientific journals covering science and medicine.

These online repositories were selected because they are common databases searched by AI and ML reviews, and they are expected to provide studies that investigate the predictive modeling of SE and SP (Table 2).

Table 2. Comparison of existing SLR for SE and SP

Citation	Focus of Survey	Number of database search	Papers Reviewed	Years Covered	Strength	Weakness
S. N. Ismail et al. (2021)	Student engagement	4 databases WoS, Science Direct, IEEE, Scopus	127 papers	Not mentioned	- Information about SE is well analysed and tabulated	- Inclusion and exclusion are not clearly stated - Year of publication covered is not mentioned - Machine learning is not discussed in detail
Tai et al. (2020b)	Student engagement	6 databases PsycINFO, ERIC, Education Source, Academic Search Complete, Scopus, WoS	250 papers	2000 to 2016	- PICO framework is used - Step of SLR is discussed in detail	- The information for SE is not well discussed - Tables and graph are not implemented to visualise the information -ML is not discussed
Subramainan and Mahmoud (2020)	Student engagement	6 databases WoS, SagePub, IEEE Xplore, Springer, ACM, Science Direct	87 papers	1984 to 2018	-The information of SE is well analysed and organized -The table and graph are fully utilized	-The process for articles filtration is not mentioned. - ML method not discussed
(Namoun & Alshanqiti, 2020)	Student performance	7 databases ACM, IEEE Xplore, Google Scholar, Science Direct, Scopus, Springer, WoS	62 papers	2010 to 2020	- PRISMA flow diagram is demonstrated - Information about SP prediction is discussed clearly - Limitations and recommendations are clearly stated - Previous literature review is compared	- Exclusion criteria are not stated - Not discussed about the unsupervised ML.
(Albreiki et al., 2021)	Student performance	6 databases ResearchGate, IEEE Xplore, Springer, ACM, Scopus, Directory of Open Access Journals	78 papers	2009 to 2021	- Inclusion and exclusion criteria clearly stated - PRISMA flow is demonstrated - Table and graph are used to demonstrate the information - Feature engineering and class balancing are discussed	- Only focused on classification ML. - Limitation is not stated
Baashar et al. (2021)	Student performance	5 databases ScienceDirect, ACM, IEEExplore, Scopus	30 papers	2015 to 2019	- PICO framework is used - PRISMA framework is demonstrated - Distribution of ML analysed	- Inclusion and exclusion criteria is not stated - ML only focus on NB, SVM, kNN, DT and ANN - Limitation and future work are not discussed
Abu Saa et al. (2019)	Student performance	5 databases ScienceDirect, EBSCO, ProQuest, JSTOR, Taylor and Francis online	36 papers	2009 to 2018	- Both clustering and classification ML are included - PRISMA framework is demonstrated	- Attributes categories is not described in detail

Step 3: Identify the Search String

This SLR implemented the search keyword formulation method by Khatibsyarbini et al. (2018) and the PICOC model by Popay et al. (2006) to identify key search terms for answering RQ. The PICOC protocol defines five key elements: population, intervention, comparison, outcome, and context, as shown in Table 3. The steps for formulating a search keyword are:

1. Determine significant search terms based on RQs using the PICOC protocol.
2. Determine equivalent words for significant search terms.
3. Determine keywords in relevant studies.
4. Use the Boolean operators 'OR' and 'AND' to link the terms.

Table 3. PICOC protocol adopted in this SLR

Population	Intervention	Comparison	Outcome	Context
HEIs student	Student engagement	Student performance	Prediction accuracy	Remote learning

Due to the search string requirements of each online repository being slightly different from each other, especially ScienceDirect and PLoSONE, three different search strings are used in this SLR. The search strings used are shown in Figure 1.

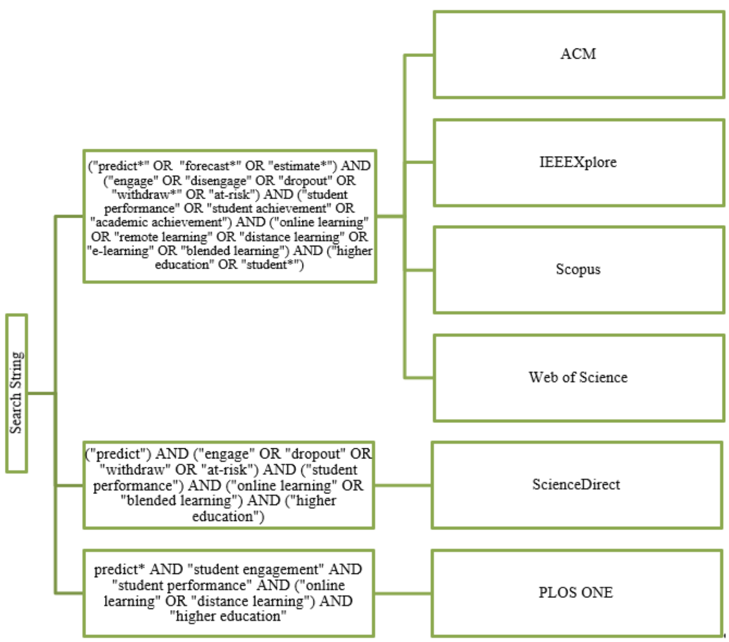

Figure 1. Search string used in this SLR

ScienceDirect and PLOS ONE use different strings because the ScienceDirect search string only allows a maximum of eight Boolean operators. Therefore, only the most significant search term is used in the search string.

Step 4: Selection of the Related Studies

PRISMA statement comprises four phases in identifying potential studies via automated and manual searches to ensure clarity and quality in the methodology (Albreiki et al., 2021; Baashar et al., 2021; Namoun & Alshanqiti, 2020). The screening phase eliminates duplicate and irrelevant studies. Qualified papers are then assessed for eligibility, resulting in the final set of studies for this SLR. Strict inclusion and exclusion criteria are applied during the screening and eligibility phases. The PRISMA flow diagram for this SLR is demonstrated in Figure 2.

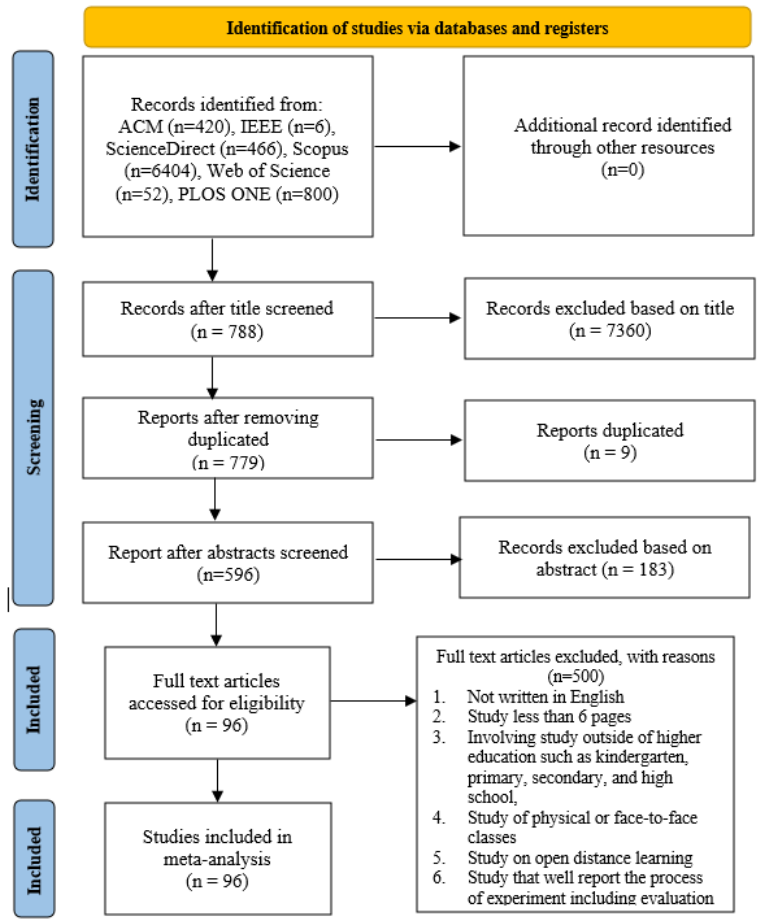

Figure 2. PRISMA flow diagram involved in this SLR

Possible sources of bias were identified and addressed systematically during the systematic review to achieve high credibility for the conclusions. The reasons for biases that might occur because of the screening process include subjectivity in the exclusion and inclusion criteria during screening, the extent of coverage achieved by the identified search terms, and the databases used in the screening process. To counteract these, the study design is categorically determined by a set of pre-definitions of inclusion and exclusion criteria. According to the PICOC framework, the search strings were built and made consistent across the databases, relevant to the specified RQs. The inclusion and exclusion criteria implemented in this research are shown in Table 4.

Table 4. Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Written in English	Not written in English
Published in years 2017 to 2024	Published earlier than 2017
Study only involves higher education	Study involves outside of higher education, such as kindergarten, primary school, and secondary school
Study involves remote learning, online learning, and distance learning	Study involves physical or face-to-face classes
Study involves formal full-time universities	Study involves open distance learning such as Udemy and Coursera

To minimize the likelihood of publication bias more, grey literature was considered where possible. Nevertheless, some sources, for example, those published in other languages, were excluded, and the effects of such exclusion on the scope of the review are considered in the following section. Manual validation was used during data extraction because errors regarding the classification of data are common with the imprecise synthesis of various studies. These measures were taken in order to achieve a high scientific output with nominal methodological ambiguity.

An initial automated search yielded 8148 articles: 420 from ACM, six from IEEE Explore, 466 from ScienceDirect, 6404 from Scopus, 52 from Web of Science, and 800 from PLOS ONE. After title screening, 7360 articles were removed, leaving 788 articles. Nine duplicates were then removed, and 183 were removed based on abstract screening. Finally, 96 articles met the including criteria and were selected for meta-analysis in this SLR from the original 8148 articles. This research is focused on higher education students since it was observed that HEI students are the most preferred sample group due to university students receiving education through distance education at a higher rate (Turan & Karabey, 2023).

STEP 5: SYNTHESIS AND EXTRACTION OF THE DATA

The principle of data synthesis is to simplify the evidence presentation for easier data extraction to answer the research questions. Using the PRISMA approach, the selected studies were thoroughly analyzed to extract the relevant information. The extracted data includes:

1. The general information, including publication year and types of articles reviewed.
2. The definition and dimension of SE.
3. The attribute categories of the educational dataset used for prediction.
4. The ML models and approaches used for predicting SE or SP.
5. The methods to improve model performance include feature selection, class balancing, and hyperparameter tuning.
6. The performance measure was implemented to evaluate the model.

The data is grouped and categorized to address each RQ and reported in the result sections.

RESULT AND DISCUSSION

This section analyses and discusses data synthesized from the SLR to answer previously mentioned RQs. It is divided into six parts: publication venues and years, definition and dimensions of SE, ML approaches for SE or SP prediction, approaches to improve model performance, and performance measurement to evaluate the model. A total of 96 studies were analyzed with publication venues, including journal articles (72 studies, 75.00%) and conferences (24 studies, 25.00%), as shown in Figure 3.

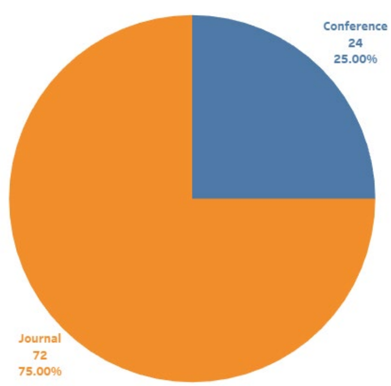

Figure 3. Type of articles reviewed

The studies observed in this SLR consist of more journal articles than conference proceedings because journal articles are more permanent than conference proceedings and are cited more by researchers (Turan & Karabey, 2023). In addition, many researchers tend to publish extended versions of the studies that are presented in the conferences as journal articles. However, in the computer science field, conferences tend to provide meaningful trends for research in this field. Therefore, high-quality conference articles are included in this SLR.

The analysis of scholarly articles from 2018 to 2024 reveals trends in SE theory, prediction of SE, prediction of SP, and prediction of SE together with SP, as shown in Figure 4. The interest in these areas surged from 2018 to 2021, dropped in 2020 and 2022, and gradually increased in 2023. There is a limited number of papers in 2024 since this review was done in May 2024.

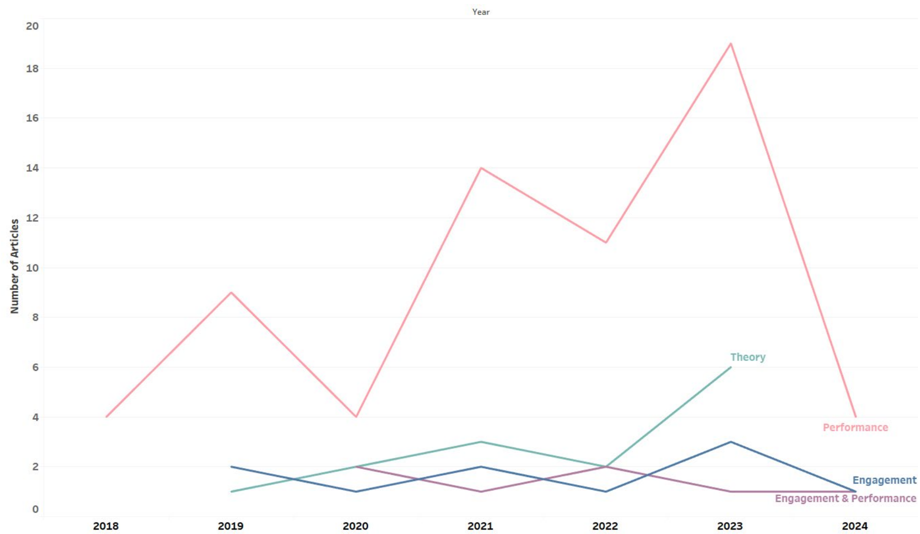

Figure 4. Distribution of articles from 2018 to 2024

The analysis of SE with SP prediction revealed fluctuation in patterns. Of note, it is quite surprising that little research has been done to predict SE compared to the vast work done in SP prediction. The increasing concern towards SE proves the need for research on more detailed prediction methods. Interestingly, despite the existence of the ML technique in the field of SE research, its application is limited when it comes to the accurate prediction of SE (S. N. Ismail et al., 2021). Identifying the factors that lie under SE can result in better educational practices and individualized assistance, which can improve academic performance and promote a positive learning environment (Wells et al., 2021).

Research Question 1: What is student engagement?

Student engagement is a critical challenge during online learning, and it needs to be facilitated in three main dimensions, which are social/emotional, cognitive, and behavioral engagement (ElSayary et al., 2022). According to Ahmadi et al. (2023), SE refers to the extent to which students invest their time, physical effort, and psychological ambition in academic tasks involving attention, curiosity, interest, optimism, and passion in educational settings (Quigley et al., 2022). SE is a complex concept with multiple dimensions (Ayouni et al., 2023; Binali et al., 2021; Gledson et al., 2021) and is essential for OL due to its predictive power for students' retention, test scores, learning, and graduation (Binali et al., 2021; Quigley et al., 2022). According to Wells et al. (2021), more active and engaged students generally perform better than less active students during online learning. Education 4.0 is intended to foster skilled persons to acquire competency in using ICT for communication and cooperation (Demong et al., 2023).

Figure 5 illustrates the categorization of SE into unidimensional, two-dimensional, three-dimensional, and four-dimensional categories. Of the 17 studies, one defined SE as behavioral and cognitive engagement. Most of the research (70.59%) employs three dimensions, and the remaining 23.53% use four dimensions, the most common being behavioral, cognitive, and emotional engagement. This is in line with SLR by S. N. Ismail et al. (2021), which adopts a broader view of self-efficacy as including behavioral, cognitive, and emotional aspects because most of the studies cited in the Fredricks et al. (2004) engagement framework. Social engagement is a commonly used dimension in the concept of SE, as found by Tai et al. (2020) and Subramainan and Mahmoud (2020).

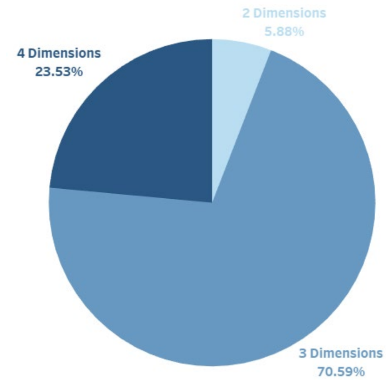

Figure 5. Distribution of articles define SE into various dimensions

Behavioral SE refers to observable actions and efforts to participate in the learning process, including class attendance, participation in academic activities and learning tasks, and use of communication tools and platforms (Ahmadi et al., 2023; Binali et al., 2021; Booth et al., 2023; Deng et al., 2020; Dragomir & Dumitru, 2023; Gledson et al., 2021; N. Luo et al., 2022; Marôco et al., 2020; Park & Yoo, 2021; Salta et al., 2021; Torres & Statti, 2023; Zafar & Nausheen, 2022). It is the physical behavior that refers to student association doing the work and following the rules during class (ElSayary et al., 2022). It relates to the total ability to finish a course, whether obstacles or barriers and to use the right amount of effort to grasp difficult ideas and acquire essential skills (Dragomir & Dumitru, 2023). It can be measured by frequency of participation in class activities, number of login clicks, attendance in virtual classrooms, amount of effort, and time spent on online learning (Ayouni et al., 2023).

Cognitive engagement refers to the amount of mental investment that is given in learning activities, motivational and self-regulated learning, which comprehends complex knowledge and skill acquisition, including reading materials and watching videos up to knowledge gain, and where the student looks for more information. (Ahmadi et al., 2023; Booth et al., 2023; Deng et al., 2020; Dubey et al., 2023; Gledson et al., 2021; N. Luo et al., 2022; Marôco et al., 2020; Park & Yoo, 2021; Torres & Statti, 2023; Zafar & Nausheen, 2022). It also refers to the active learning process, the most essential learning form during online learning (ElSayary et al., 2022). Cognitive engagement encompasses everything from the ability to concentrate to the use of advanced learning techniques, including memory recall, acquisition of knowledge, and a thorough comprehension or mastery of the work at hand (Booth et al., 2023; Dragomir & Dumitru, 2023). It can be assessed through course time and active involvement in supplementary quizzes, which refer to students' eagerness and capability to do the learning task (Ayouni et al., 2023). It is important for student achievement during distance learning since it promotes the students' behavioral engagement (Hsueh et al., 2022).

Emotional engagement is students' perception of instructors, peers, academic environment, and educational institution (Booth et al., 2023; Deng et al., 2020; Dubey et al., 2023; Marôco et al., 2020; Park & Yoo, 2021; Salta et al., 2021; Torres & Statti, 2023). The construct can be measured by curiosity, enjoyment, and belongingness (Ahmadi et al., 2023; Dubey et al., 2023; Marôco et al., 2020; Torres &

Statti, 2023). Furthermore, it can be shown by the actions that the student associates with learning, such as excitement, interest, and motivation (ElSayary et al., 2022). It includes such approaches as increasing the enticement value of learning resources (Quigley et al., 2022). Evaluation and assessment of the extent and nature of students' emotional bonding, especially in the context of OL, is challenging, even though it is a fundamental element of effective OL (Ayouni et al., 2021).

In addition, social engagement should be incorporated into OL to encourage students' participation and social interactions within the OL environment through online discussion forums and sharing of learning materials (Ayouni et al., 2023; Binali et al., 2021; Deng et al., 2020). Additionally, it included the propensity for creating and maintaining relationships in terms of OL (Binali et al., 2021; Deng et al., 2020; Quigley et al., 2022). The metric can be quantified through the frequency of interactions, including conversations, chats, and emails, with both peers and teachers (Ayouni et al., 2023). Social engagement greatly influences online interaction and can be the main driver of engagement (Ayouni et al., 2023).

SE incorporates behavior, cognition, emotions, and social interactions, which have different functions in learning. Behavioral engagement is composed of tangible behaviors such as active participation and putting effort into academic tasks. Cognitive engagement involves how deeply students are engaged intellectually, while emotional engagement requires students' emotional connection with their educational experience. Social engagement focuses on the importance of interactions and relationships within a learning environment. Understanding SE is important for enhancing student retention, performance, and overall achievements, particularly in an OL context.

Research Question 2: What is the Popular Data Used to Predict Student Engagement or Student Performance?

It is important to select appropriate attributes for the prediction of SE and SP to ensure an accurate predictive model. Figure 6 indicates that 92.41% (73 research) use LMS logfile behavioral attributes to predict SE or SP. J. Chen et al. (2019) and Wells et al. (2021) determine that learning behavior can also be considered one of the most useful indicative measures of SE in educational activities. The collection and analysis of student learning behavior can give informed decisions for the syllabus designer and lecturer on the subject design and provide more informed ways to deliver the learning material (Wells et al., 2021). Student behavior measures include attendance, resource views, course material downloads, assignment submissions, quiz completions, video watching time, assignment time, quiz time, forum views, and forum participation. Student behavior data are focused, while LMS is a commonly used platform to improve learning processes and create engaging and effective learning (Palanci et al., 2024).

Furthermore, 41.77% (33 papers) employ academic variables, like mid-term scores, assessment scores, grades, and performance, for predicting SE or SP. The demographic category attributes such as gender, age, parent's income, address, and scholarship are utilized in 36.71% (29 studies). The application of academic background and psychological qualities is restricted, with only 11.39% (9 research) and 1.27% (1 study). Currently, academic backgrounds and psychology are not considered when predicting SE, while the implementation of academic measures is very limited. Demong et al. (2023) suggest data can be widened to include other personality dimensions as well as the digital competency of the student. Unlike Abu Saa et al. (2019), there are more e-learning activities than student academic attributes in this review because of the emphasis on OL.

The SE dataset is a valuable attribute for SP and SE prediction since students' performance can be improved by increasing the SE level in the learning process, especially during OL (Bernacki et al., 2020). SE is not a unidimensional construct but is made up of behavioral, cognitive, emotional, and social components (Gledson et al., 2021). However, most of the research carried out on SE tends to present it in an overgeneral manner with little or no regard for the multidimensionality of the concept.

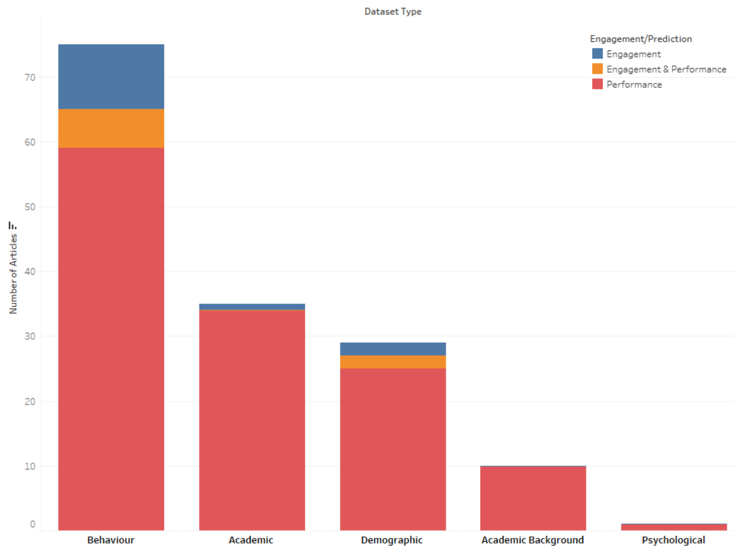

Figure 6. Attribute category used for SE and SP predictions

Ultimately, the most crucial factor in enhancing SP and students' performance is the SE's accurate prediction. The broad usage of LMS logfile behavioral attributes (92.41%) emphasizes their importance in educational research. According to Palanci et al. (2024), LMS is one of the most preferred platforms in educational studies, and learning behavior is found to be the most examined, and in this educational research, the log data are mainly utilized. Some important attributes include attendance, resource views, and time allocation to provide SE insight and affect performance.

Although academic, demographic, and psychological attributes contribute to predictive models, their implementation is limited. It is important to understand that SE is a complex concept that includes behavior, cognition, emotions, and social dimensions. Meanwhile, it is necessary to mention that such concerns as violence and discrimination against women are still topical and affect modern culture independently of the environment. Education and government undertakings to combat gender concerns are difficult to harmonize, lacking or, at times, even absent (Yañez et al., 2023). Therefore, it is crucial to acknowledge that SE is a multifaceted process that informs efficient paraphrasing of estimates in terms of educational results that enable precise interventions.

Research Question 3: What are the Popular Machine Learning Methods Used to Predict Student Engagement or Student Performance?

ML methods are categorized into unsupervised, which includes clustering techniques, and supervised ML, which covers classification and regression techniques. Figures 7 and 8 show the distribution of ML used in educational research.

Figure 7 shows that ML methods are mainly divided into classification, clustering, and regression. Classification is the most popular ML method implemented in the SP and SE research, used in 88.60% (70 papers), with 78.48% (62 papers) relying solely on classification, and 7.59% (6 papers) and 2.53% (2 papers) combining with clustering and regression, respectively. Clustering follows at 15.19% (12 papers), with 5.06% (4 papers) implementing only clustering, 2.53% (2 papers) combining with regression, and 7.59% (6 papers) implementing classification. Lastly, regression is least used, with only 12.65% (10 papers), 7.59% (6 papers) use it alone, and 2.53% (2 papers) each with clustering and classification, respectively. Regression is not discussed in detail for this review due to its limited application to educational outcome prediction, especially for SE.

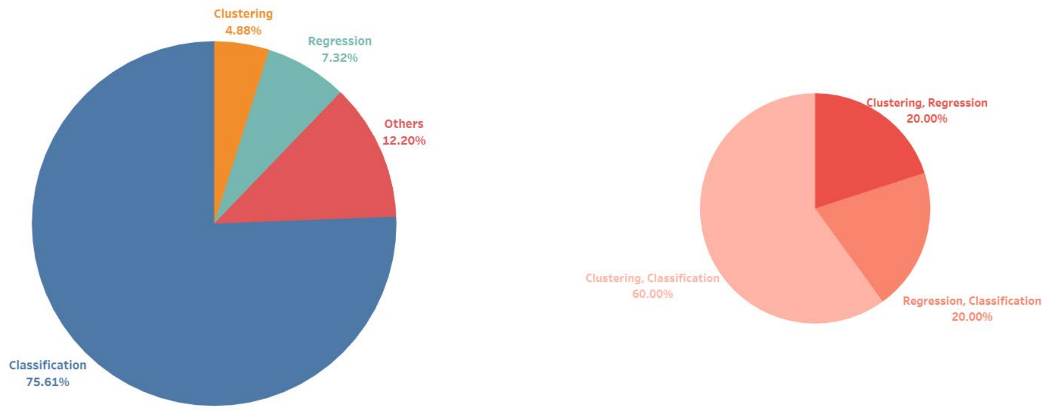

Figure 7. ML methods used in SE and SP prediction

Figure 8 shows that this review categorizes the referred studies into the prediction SE, SP, SE, and SP. Most of the SP predictions are implementing classification ML models. However, there are limited studies of SE implementing classification ML, where clustering ML is commonly implemented to categorize the level of SE. This finding is similar to the research of Helal et al. (2018) and Y. Yang et al. (2020), who mentioned that classification is crucial for SP prediction and clustering is key for indicating the students' learning behavior or SE. Classification is normally preferred compared to clustering and regression for the prediction of SE and SP due to its ability to provide clear, actionable insight into student categories. This is because the classification ML model can effectively categorize students into different performance levels and engagement levels.

Furthermore, classification helps identify the significant factors that influence student performance, which can help the institutions tailor educational strategies (Ghosh et al., 2022). The lack of clustering implementation in educational research is because clustering does not provide clear-cut partitioning of SP. Hence, the result is less insightful, useful, and actionable in an educational context. On the other hand, clustering ML is an unsupervised learning technique that clusters the unlabelled data by simply looking at a set of objects and establishing a group of similar traits (Yağcı, 2022).

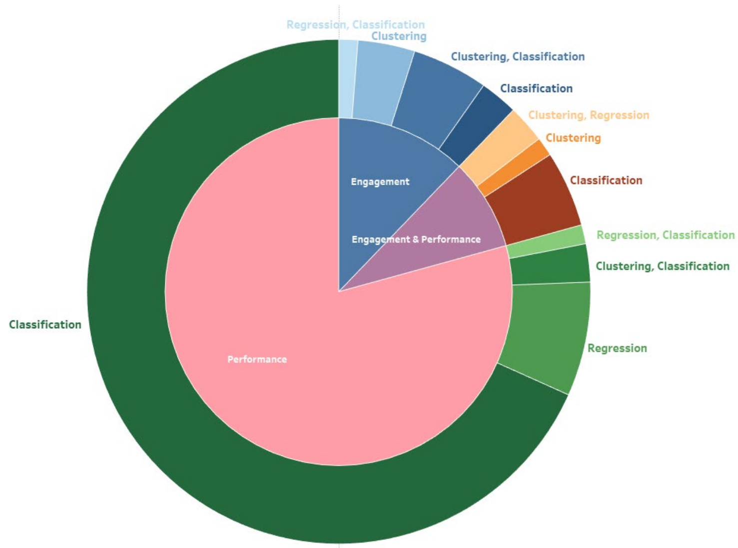

Figure 8. Distribution of ML implemented for SE and SP prediction

Classification organizes samples into classes based on similar traits (Helal et al., 2018), using labeled data samples to predict tasks (Tomasevic et al., 2020). Meanwhile, the clustering ML method is an unsupervised learning approach that clusters the unlabelled data by finding and grouping objects in the same group based on similar traits (Hasan et al., 2020). Classification is preferred for the prediction of SP since it has clear academic result labels, while SE lacks such labels, making clustering useful for automated labeling (Nguyen et al., 2018). However, classification is getting important and momentum to better understand and optimize the learning process and environments (Tomasevic et al., 2020). Tomasevic et al. (2020) and Hasan et al. (2020) suggest clustering should precede classification for effective data tagging.

Figure 9 highlights the top 10 popular ML methods for classification. RF is the most popular used in 62.86% (44 papers) of research. RF, SVM, and DT are used in 42.86% (30 papers) and 41.43% (29 papers), respectively. Last, NB and Logistic Regression (LR) are also common, implemented in 38.57% (27 papers) and 34.29% (24 papers) of the research. DT, NB, and LR are XAI, while RF and SVM are non-XAI. The RF and SVM are popular due to their high prediction effectiveness, which can provide higher accuracy. However, they do not provide explainable outcomes of the results obtained.

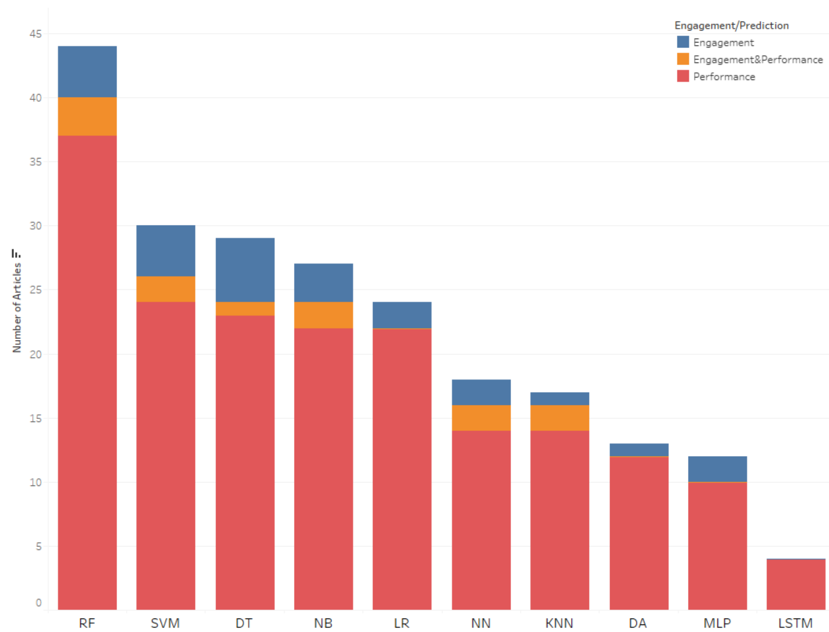

Figure 9. Top 10 Popular ML for Classification

RF is an ensemble approach that combines multiple DTs and finalizes the decisions based on majority voting mechanism (Al-Shabandar et al., 2019; Gkontzis et al., 2018; Guo et al., 2020; Saleem et al., 2021). The main reason for constructing multiple DTs with different combinations is to improve the predictive accuracy (Saleem et al., 2021). RF is a flexible and user-friendly ML algorithm that is easy to use for classification (Hussain et al., 2018). It is also stable, reduces bias by combining outputs from multiple DTs, and uses random samples to build trees (Orji & Vassileva, 2020). RF is less computationally consuming and can prevent overfitting issues.

Meanwhile, it is not sensitive to noise and can handle high-dimensional data. RF is popular due to its ability to provide a high accuracy rate, which is up to 98.25% accuracy when predicting the SP using students' learning behavior datasets (J. Chen et al., 2024). Due to its robustness and overfitting advantages, RF is appropriate for dealing with diverse educational datasets such as dropout prediction. However, when the model complexity increases, it will become challenging for decision-making and complicate communication of results to stakeholders (Manzali et al., 2024).

NB is a probabilistic approach that uses the attributes of the data set and the Bayesian theorem to derive the probability of each class (Guo et al., 2020; Helal et al., 2018; Tomasevic et al., 2020). Inverse probability NB can infer unknown quantities and make predictions from data as in the inspection of NB (Tomasevic et al., 2020). NB can work with big data, is simple to implement, and allows for the analysis and prediction of high-dimensional data (Saleem et al., 2021). It is also a chiefly ranker algorithm and can work in binary, multiclass classification without compromising runtime performance due to its scalability (Tomasevic et al., 2020). Naïve Bayes can provide high accuracy rates for predicting student performance and provide prediction results within a short time, irrespective of the large educational dataset used (Akanbi, 2023). However, NB assumes that all the attributes are independent, which might not be the case when using attributes in real-world education contexts (Buhori et al., 2024). Furthermore, NB will face classification issues when dealing with datasets with complicated relationships (Manzali et al., 2024).

SVM is a nonlinear function approach that utilizes hyperplanes to separate the attribute spaces (Gkontzis et al., 2018; Ma et al., 2018). It represents feature vectors as points in multidimensional space, mapping them to different classes that are divided by a maximally wide plane (Zheng et al., 2020). SVM is advantageous because it is not affected by the dimensionality of the sample; even small training data can maintain a strong adaptability to the test sample (Ma et al., 2018). This technique is appropriate when the dataset is small, nonlinear, and has high dimensions (Zheng et al., 2020). SVM is a powerful and flexible modeling technique used for classification (Damuluri et al., 2019). SVM works nicely for problems with complex patterns in the results of the students' evaluations. However, SVM heavily fails for the imbalanced datasets, which leads to bad predicting results for the minority class (Anisa et al., 2024).

DT is useful for analyzing data by splitting it into a tree-based structure. DTs build and develop the classification trees per some logical rules, while the leaf nodes hold the results and decision nodes provide directions (Saleem et al., 2021). Even though DTs perform well with small samples, they might not fully represent the overall trends (Cagliero et al., 2021). They are widely used because they are easy to implement and useful for classification and prediction (Saleem et al., 2021). The interpretability of DT is relatively high, making it easy to represent a graphical design, enabling the lecturers to grasp the factors in students' performance and make decisions on time to enhance student performance (Muriana et al., 2022). However, DT can easily act as an overfitting model and make it less reliable toward a limited or biased set of data found in the educational sector (Nagarajan et al., 2024).

LR is a supervised ML technique that uses generalized logistic regression analysis to estimate the probability of an event occurring according to the regression coefficient of one or more characteristic variables (Kabathova & Drlik, 2021; Zheng et al., 2020). It is normally implemented to classify discrete variables (Zou et al., 2020). The advantage of LR is that it is simple and efficient in calculating the future outcome using a linear equation (Karalar et al., 2021). LR can perform well without complex feature engineering (Li et al., 2020). The LR's high interpretability can give a good indication of the importance of the effect of the different factors on student performance and help the lecturers to define important variables such as demographic and academic background (Kurniadi et al., 2023). In addition, insight into the learning environment can act as an early warning sign that can help identify at-risk learners and provide timely support (Sagala et al., 2022). LR assumes a linear relationship between independent variables and the log of the dependent variables, which may not be appropriate in many complex educational data sets (Sagala et al., 2022).

Figure 10 shows the clustering ML techniques used in 31 studies. K-means clustering is the most popular method, and it was implemented in 83.33% (10 papers) of the studies. Other clustering methods are less common, such as agglomerative (3 papers), spectral (3 papers), DBSCAN (2 papers), and hierarchical and Birch (1 paper each). Therefore, only the K-Means clustering is focused in this review to answer the RQ3.

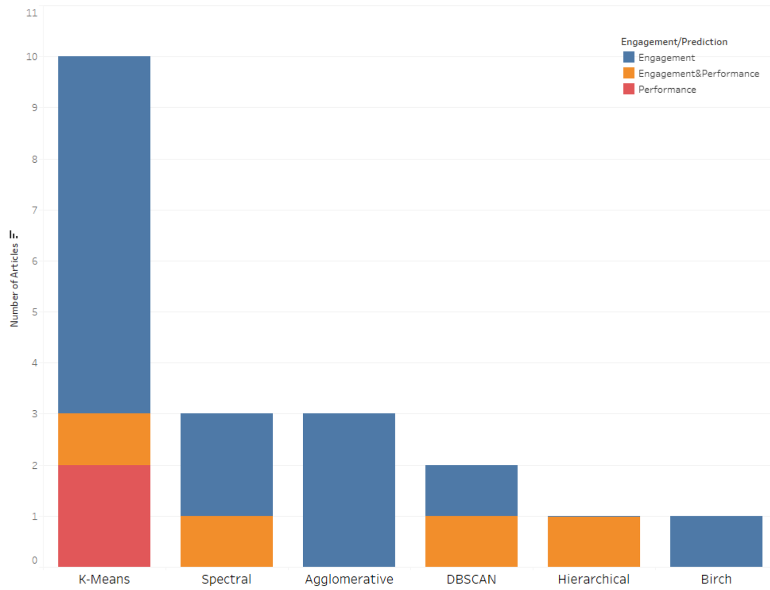

Figure 10. ML clustering methods

K-means clustering is an unsupervised learning algorithm that looks at groups in unlabelled data by dividing observations into *k* clusters, each belonging to the cluster center with the nearest mean (Hussain et al., 2018; J. Luo & Wang, 2020). This method indicates that when a dataset consists of *n* samples, and each is a vector with *d* dimensions, samples can be aggregated into *K* clusters (Lu et al., 2021). K-means clustering is favored for its ability to produce clusters with significant differences (Sheik Abdullah et al., 2021). K-means clustering is straightforward to understand and implement. This makes it accessible for lecturers and researchers to interpret the hidden pattern of the educational dataset. However, k-means calculate centroid averages, which might not always indicate true student performance, especially in split score datasets (Badhera et al., 2022).

Therefore, predicting SE by applying ML models to improve educational results is crucial. Classification techniques prevail and are implemented in 82.69% of works since collecting labeled data allows for high accuracy in predicting SP. However, clustering methods, applied in 19.87% of cases, could be useful for predicting SE since this area is frequently characterized by the lack of clearly defined data sets. SE prediction with classification can enhance the learning process and the student's performance, especially when predictors are well chosen. However, using SE prediction with classification is still limited since labeled sample data are rare. The interdisciplinary collaboration with educational experts will collect their knowledge by taking psychological components on differentiating and labeling student engagement into distinct levels that can be used to train the classification model.

RF is the most popular classifier, followed by SVM, DT, NB, and LR for educational prediction, while K-means clustering is preferred for grouping unlabelled data effectively. The classification approaches of DT, NB, and LR XAI should focus on educational research due to their explainability and transparency. In the future, feature engineering needs to be studied for educational datasets to standardize and reduce the complexity of the dataset to improve the performance of XAI while enhancing the transparency and explainability of the model. Demong et al. (2023) mentioned that different classification algorithms should be applied to proactively predict student engagement levels in the future. Hybrid machine learning is also possible if different algorithms can be integrated to eliminate their respective drawbacks and improve the prediction capacity.

Research Question 4: How to improve the performance of the prediction model?

The main goal of developing a predictive model is to achieve accuracy. The approaches that can be implemented to improve predictive accuracy can be divided into improving input data quality and optimizing the algorithm (Ali et al., 2023; L.-q. Chen et al., 2021). Feature selection (FS) and resampling (RS) approaches can be implemented to improve the quality of data (L.-q. Chen et al., 2021; Rozi et al., 2023). Meanwhile, hyperparameter tuning (HT) is utilized to advance the ML algorithm (Ali et al., 2023).

The reviewed papers use different approaches to improve the prediction model performance, as shown in Figure 11. FS is the most common approach, which is implemented in 29.11% (23 papers), while RS strategies are used in 18.99% (15 papers) and HT in 18.99% (15 papers). Despite being the second most popular method, RS is rarely used for predicting SE. Only two of the papers on SE prediction implement the RS technique. However, in a world where education is already characterized by unfairness or bias towards certain communities, resulting in dataset imbalance issues, the model excessively emphasizes unique characteristics of over-represented groups, resulting in more accurate predictions for that population but less accurate predictions for the under-represented group (Jamaluddin & Mahat, 2021).

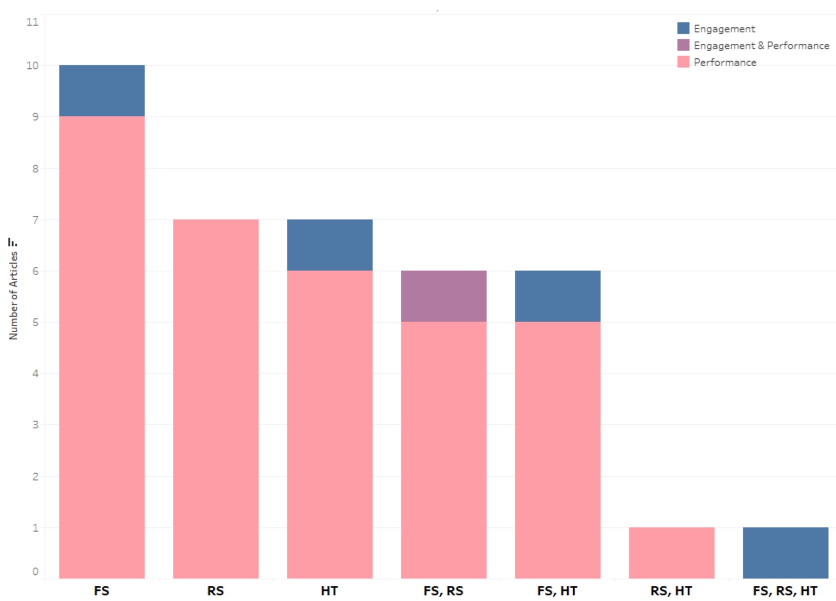

Figure 11. ML enhancement approaches

Ramaswami et al. (2020) highlight that utilizing all the features leads to overfitting and reduces the generalization of the model. FS, which removes redundant or irrelevant features, is essential to prevent the curse of dimensionality and improve model accuracy (Al-Shabandar et al., 2019; Ramaswami et al., 2020; Xie et al., 2021). FS is important for XAI to provide clearer insight into the impact of each attribute on prediction outcomes to make the model more straightforward and transparent (Zacharias et al., 2022). FS can be divided into filter-based and wrapper-based (Yamasari et al., 2020). Filter-Based Feature Selection (FBFS) uses statistical and probabilistic measures to evaluate the relevance of each feature, independent of the ML algorithm (Ma et al., 2018; Yamasari et al., 2020), whereas Wrapper-Based Feature Selection (WBFS) uses a classifier to assess the importance of a feature subset based on its ability to improve prediction accuracy (Yamasari et al., 2020). However, WBFS is more time-consuming since it evaluates each feature with the classifier. FBFS, including Information Gain (IG), Principal Component Analysis (PCA), correlation-based FS, and Chi-square FS, are more preferred in educational prediction due to their transparency and explainability.

Class imbalance is one of the challenges in educational research that affects the performance of conventional classification algorithms. RS, which includes oversampling, undersampling, and hybrid techniques, can be used to balance class distribution in training data (Hassan et al., 2021; Jamaluddin & Mahat, 2021). Oversampling duplicates minority samples to balance with majority samples, which include Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), and Random Oversampling (ROS) (Hassan et al., 2021). However, it might cause overfitting due to the generation of new synthetic samples (Hassan et al., 2021). Under-sampling like Tomek Links (TL), Edited Nearest Neighbours (ENN), and Random Under-sampling (RUS) (Hassan et al., 2021), removing samples from the majority class of the dataset might cause the loss of information. Hybrid sampling combines both oversampling and undersampling, like SMOTE-TL, SMOTE-ENN, SMOTEBoost, and RUSBoost. SMOTE is preferred for educational prediction because of its fast and reliable sample generation, but it disregards the distribution of the minority classes and possibly hidden noisy data (Palli et al., 2024).

Hyperparameter tuning is conducted to determine the best parameter for the learning algorithm to achieve the highest prediction accuracy (Al-Shabandar et al., 2019; Karalar et al., 2021). Grid Search, Random Search, and Tree-structured Parzen Estimator (TPE) can be implemented to optimize hyperparameters for each model. Hassan et al. (2021) suggest that the implementation of HT will improve the predictive model. Grid Search systematically uses different parameter combinations to optimize the model's performance and make it user-friendly and effective (Ma et al., 2018). Random Search carries out a comprehensive parallel search across specified parameter values (Riestra-González et al., 2021). TPE implements a sequential model-based optimization approach to estimate hyperparameter performance based on previous measurements and select new ones to test accordingly.

In conclusion, FS and RS need to be implemented to improve SE prediction accuracy in terms of data quality. FS that eliminates irrelevant or redundant attributes can reduce overfitting issues and improve model accuracy. RS overcomes the data imbalance challenges while ensuring more fair and accurate prediction across various is not widely implemented for student engagement prediction. The implementation of FS and RS is crucial to enhance SE prediction and foster better educational practices.

Research Question 5: How to Measure the Prediction Performance of the Machine Learning Algorithm?

Performance measures (PM) are crucial for evaluating and comparing the effectiveness of predictive models. Selecting and implementing appropriate PMs is essential for assessing model performance. This review will discuss popular PMs used by researchers for clustering and classification.

Figure 12 shows common PM for clustering methods. Notably, 50% (6 studies) do not employ any PM, while only six papers apply PM to assess their models. Of the six papers, four use the silhouette coefficient as the best measure for clustering models, and hence, it has become one of the most common.

The silhouette coefficient is a common method for determining the optimal number of clusters in a dataset (J. Luo & Wang, 2020). It is calculated using the average intra-cluster distance (*a*) and average nearest neighbor (*b*) for each sample. The formula for the silhouette coefficient is demonstrated as follows:

$$\frac{(b-a)}{\max(a,b)}$$

While 1 is the best value and -1 is the worst value, the higher values indicate that the object is well-matched to its cluster and poorly matched with the neighboring clusters (J. Luo & Wang, 2020).

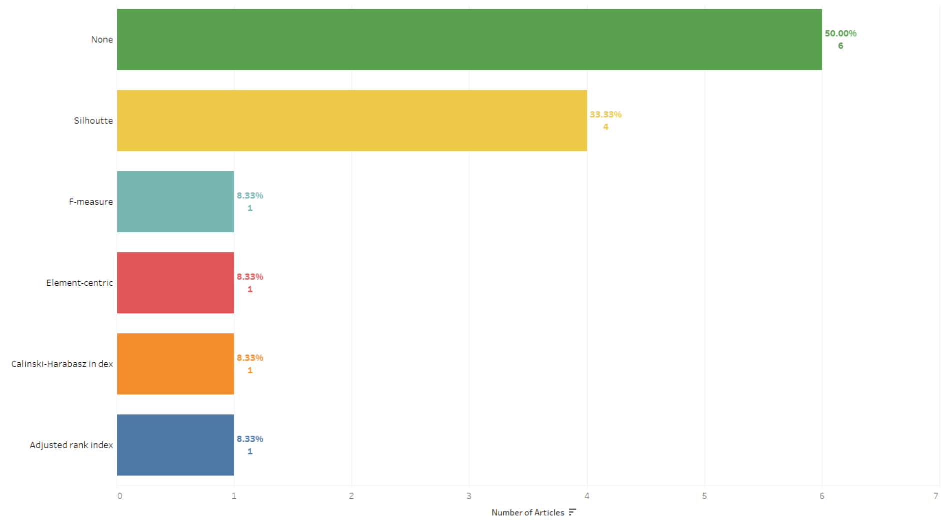

Figure 12. Performance measure for clustering method

Figure 13 shows the top five PM for evaluating classification models. Accuracy is the most popular, and it is used in 75.71% (53 papers) of the studies. Recall is used in 65.71% (46 papers), precision in 58.57% (41 papers), F1-score in 57.14% (40 papers), and Area Under the Curve (AUC) in 31.43% (22 papers). Among these, Accuracy, Recall, Precision, and F1-score are used in over half of the papers, which aligns with the findings of N. Ismail and Yusof (2022).

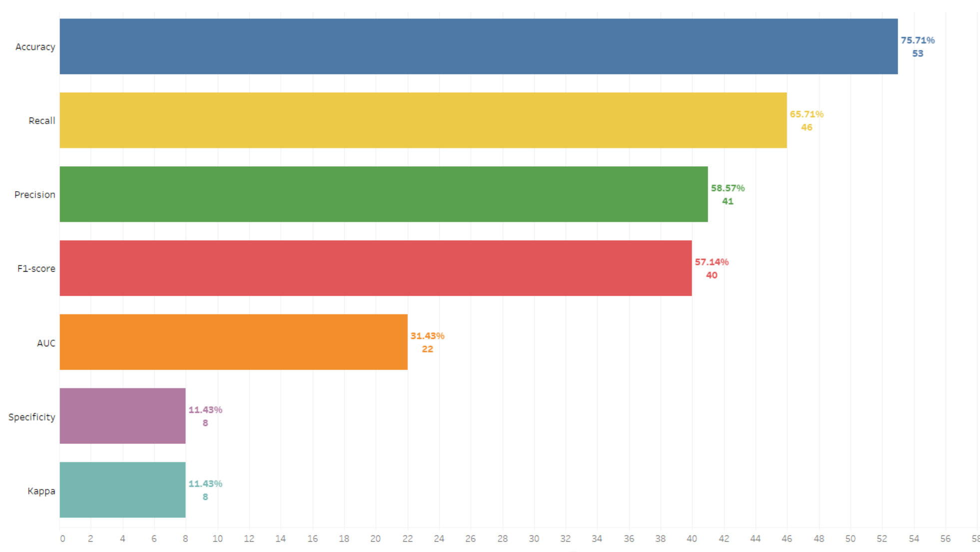

Figure 13. Top 5 performance measure for classification methods

Accuracy is the proportion of total numbers that are predicted correctly (Zheng et al., 2020), which is the number of samples that are classified correctly and divided by the total number of items (Damuluri et al., 2019). Accuracy is calculated using the equation:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Recall is one such validation criterion, also known as sensitivity, used to evaluate algorithm performance (Ayouni et al., 2021; Saleem et al., 2021). Recall represents how many samples are correctly predicted and shows how complete a prediction result is with respect to the fraction of correct classifications of all positive samples (Helal et al., 2018; Zheng et al., 2020). The recall formula can be expressed as follows:

$$Recall = \frac{TP}{TP + TN}$$

Precision is the accuracy of positive predictions that shows the predictive power of a classifier (Ayouni et al., 2021; Zheng et al., 2020). It is the fraction of true positive samples of all classified positive examples reflecting the exactness of predictions (Helal et al., 2018). Precision is calculated as the ratio of true positive values to the sum of all positive predictions, as shown in the calculation (Saleem et al., 2021):

$$Precision = \frac{TP}{TP + FP}$$

F1-score is a common multiclass metric used to assess the classifier's performance (Saleem et al., 2021). It is the harmonic mean of precision and recall, indicating overall performance and the balance between precision and recall (Helal et al., 2018; Karalar et al., 2021; Saleem et al., 2021). F1-score indicated the classifier performance that was evaluated comprehensively after reconciling the recall and precision rates (Zheng et al., 2020). F1-score can be calculated by using the formula as follows:

$$F1 - score = \frac{2 * Recall * Precision}{Recall + Precision}$$

where TP, TN, FP, and FN represent the True Positive, True Negative, False Positive, and False Negative, respectively.

To conclude, the silhouette coefficient is the popular PM used to evaluate the clustering methods, while accuracy, recall, precision, and F-measure are popular for the evaluation of classification methods.

Limitations and Threats to Validity

Like most previous reviews, this SLR has limitations that reduce its validity. One major limitation is that in identifying primary studies, there might be an insufficiency and incompleteness in identifying such studies arising from difficulty in developing and using search terms across various databases. To do this, the PICOC model was applied to enhance the quality of the search strings by aligning them to the appropriate level of PICOC components and the RQs. While the process of database identification and search methods was outlined, it is agreed that some studies on the prediction of SP and SE might be missing, especially local research conducted and published in languages other than English. One of the weaknesses is that data synthesis and extraction are often imprecise, and it is possible to misclassify data, distorting the review's validity. In order to overcome this, the standard PRISMA framework was followed to manually validate the correctness and accuracy of the information extracted and classified based on the inclusion and exclusion criteria set. However, there are some limitations in the present efforts, including excluding any quantitative or qualitative work published in non-English.

Conclusion

This SLR aims to assess current trends in predicting SE and SP. Five research questions guide this SLR, focusing on finding, classifying, evaluating, and understanding primary studies. The study aims to identify areas for improvement through a comprehensive assessment of relevant research in SE and SP predictions. In the process, relevant primary studies are selected and appraised, and data is extracted from them. The outcomes are given in the form of tables and figures to raise comprehension of SE and SP prediction. Key insights include:

(a) SE is made up of four main dimensions: behavioral, cognitive, emotional, and social.
(b) SE attributes are the popular dataset that is used for both SE and SP prediction.
(c) SP prediction commonly uses classification methods, while SE prediction often relies on clustering methods.
(d) RF, NB, SVM, DT, and LR are the popular classification algorithms, with DT, NB, and LR being known for their transparent and explainable capabilities.
(e) K-means clustering is widely used in clustering methods.
(f) Additional methods like FS, RS, and HT are implemented to enhance predictive model performance and the explainability and transparency of predictions. However, their application for SE and SP prediction remains limited. There is a need for methods that can handle both dimensionality and class imbalance while keeping the original data's meaning intact.
(g) The silhouette coefficient is the popular PM for the evaluation of clustering methods, while accuracy, recall, precision, and f-measure are commonly used to evaluate classification methods.

This research becomes important to the future of educational planning and development, especially when universities and other learning institutions tap data insight to improve performance. Despite the increasing focus on applying machine-learning approaches within education, there is limited research employing this approach to investigate student engagement since defining and modeling this complex construct present difficulties. Engagement labeling can be fine-tuned with the help of other scholars of interdisciplinary fields in collaboration with educational expertise and psychological knowledge. Further, feature engineering plays a significant role in normalizing and simplifying datasets to improve XAI models' efficiency, intelligibility, and transparency. Feature engineering also needs to be tackled to solve the class imbalance issue derived from most educational datasets, where some classes are under-represented. The current study paves the way for more efficient and large-scale applications of machine learning in analyzing student engagement by evaluating these challenges through interdisciplinary cooperation and sophisticated data preprocessing methods.

ACKNOWLEDGMENTS

The authors thank the Universiti Teknologi Malaysia (UTM) for giving them the opportunity and providing support to carry out this research. Besides, the author would like to sincerely thank the Pusat Pengurusan Data Universiti (PPDU) UTM and the FC Academic Office UTM for their invaluable support and prompt approval of student data for this research.

REFERENCES

Abu Saa, A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques. *Technology, Knowledge and Learning*, *24*, 567-598. https://doi.org/10.1007/s10758-019-09408-7

Agrawal, R., & Sharma, K. (2024). An extensive review on significance of explainable artificial intelligence models in discrete domains for informed decisions making. *Revue d'Intelligence Artificielle*, *38*(3), 957-968. https://doi.org/10.18280/ria.380321

Ahmadi, G., Mohammadi, A., Asadzandi, S., Shah, M., & Mojtahedzadeh, R. (2023). What are the indicators of student engagement in learning management systems? A systematized review of the literature. *International Review of Research in Open and Distributed Learning*, *24*(1), 117-136. https://doi.org/10.19173/irrodl.v24i1.6453

Akanbi, O. B. (2023). Application of Naive Bayes to students' performance classification. *Asian Journal of Probability and Statistics*, *25*(1), 35-47. https://doi.org/10.9734/ajpas/2023/v25i1536

Alamri, R., & Alharbi, B. (2021). Explainable student performance prediction models: A systematic review. *IEEE Access*, *9*, 33132-33143. https://doi.org/10.1109/ACCESS.2021.3061368

Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences*, *11*(9), 552. https://doi.org/10.3390/educsci11090552

Ali, Y. A., Awwad, E. M., Al-Razgan, M., & Maarouf, A. (2023). Hyperparameter search for machine learning algorithms for optimizing the computational complexity. *Processes*, *11*(2), 349. https://doi.org/10.3390/pr11020349

Al-Shabandar, R., Hussain, A. J., Liatsis, P., & Keight, R. (2019). Detecting at-risk students with early interventions using machine learning techniques. *IEEE Access*, *7*, 149464-149478. https://doi.org/10.1109/ACCESS.2019.2943351

Anisa, Y., Erika, W., & Azmi, F. (2024). Enhancing student performance prediction using a combined SVM-radial basis function approach. *International Journal of Innovative Research in Computer Science & Technology*, *12*(3), 1-5. https://doi.org/10.55524/ijircst.2024.12.3.1

Ayouni, S., Hajjej, F., Maddeh, M., & Al-Otaibi, S. (2021). A new ML-based approach to enhance student engagement in online environment. *PLoS ONE*, *16*(11), e0258788. https://doi.org/10.1371/journal.pone.0258788

Ayouni, S., Hajjej, F., Maddeh, M., & Al-Otaibi, S. (2023). Innovations of materials for student engagement in online environment: An ontology. *Materials Today: Proceedings*, *81*, 470-476. https://doi.org/10.1016/j.matpr.2021.03.636

Baashar, Y., Alkawsi, G., Ali, N. a., Alhussian, H., & Bahbouh, H. T. (2021, July). Predicting student's performance using machine learning methods: A systematic literature review. *Proceedings of the International Conference on Computer & Information Sciences, Kuching, Malaysia*, 357-362. https://doi.org/10.1109/ICCOINS49721.2021.9497185

Badhera, U., Verma, A., & Nahar, P. (2022). Applicability of K-medoids and K-means algorithms for segmenting students based on their scholastic performance. *Journal of Statistics and Management Systems*, *25*(7), 1621-1632. https://doi.org/10.1080/09720510.2022.2130566

Bagunaid, W., Chilamkurti, N., Shahraki, A. S., & Bamashmos, S. (2024). Visual data and pattern analysis for smart education: A robust DRL-based early warning system for student performance prediction. *Future Internet*, *16*(6), 206. https://doi.org/10.3390/fi16060206

Benabbes, K., Housni, K., Hmedna, B., Zellou, A., & Mezouary, A. E. (2023). A new hybrid approach to detect and track learner's engagement in e-learning. *IEEE Access*, *11*, 70912-70929. https://doi.org/10.1109/ACCESS.2023.3293827

Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting achievement and providing support before STEM majors begin to fail. *Computers & Education*, *158*, 103999. https://doi.org/10.1016/j.compedu.2020.103999

Binali, T., Tsai, C.-C., & Chang, H.-Y. (2021). University students' profiles of online learning and their relation to online metacognitive regulation and internet-specific epistemic justification. *Computers & Education*, *175*, 104315. https://doi.org/10.1016/j.compedu.2021.104315

Booth, B. M., Bosch, N., & D'Mello, S. K. (2023). Engagement detection and its applications in learning: A tutorial and selective review. *Proceedings of the IEEE*, *111*(10), 1398-1422 https://doi.org/10.1109/JPROC.2023.3309560

Buhori, I., Marjito, M., Supriatna, A., & Wahyudi, H. (2024). Implementation of the Naïve Bayes method with approach holistic in predicting student achievement: At one of the vocational high schools in Bandung. *Accounting and Management Journal*, *4*(1), 59-66. https://doi.org/10.55208/aj.v4i1.139

Cagliero, L., Canale, L., Farinetti, L., Baralis, E., & Venuto, E. (2021). Predicting student academic performance by means of associative classification. *Applied Sciences*, *11*(4), 1420. https://doi.org/10.3390/app11041420

Chen, J., Feng, J., Sun, X., Wu, N., Yang, Z., & Chen, S. (2019). MOOC dropout prediction using a hybrid algorithm based on decision tree and extreme learning machine. *Mathematical Problems in Engineering*, *2019*(1), 8404653. https://doi.org/10.1155/2019/8404653

Chen, J., Zhou, X., Yao, J., & Tang, S.-K. (2024, March). Evaluation of student performance based on learning behavior with Random Forest Model. *Proceedings of the 13th International Conference on Educational and Information Technology, Chengdu, China*, 266-272. https://doi.org/10.1109/ICEIT61397.2024.10540693

Chen, L.-q., Wu, M.-t., Pan, L.-f., & Zheng, R.-b. (2021). Grade prediction in blended learning using multisource data. *Scientific Programming*, *2021*(1), Article 4513610. https://doi.org/10.1155/2021/4513610

Ch'ng, C. K. (2024). Hybrid machine learning approach for predicting E-wallet adoption among higher education students in Malaysia. *Journal of Information and Communication Technology*, *23*(2), 177-210. https://doi.org/10.32890/jict2024.23.2.2

Damuluri, S., Ahmadi, P., & Islam, K. (2019). A study of several classification algorithms to predict students' learning performance. Paper presented at 2019 ASEE annual conference exposition. Tampa, Florida.

Demong, N. A. R., Shahrom, M., Abdul Rahim, R., Omar, E. N., & Yahya, M. (2023). Personalized recommendation classification model of students' social well-being based on personality trait determinants using machine learning algorithms. *Journal of Information and Communication Technology*, *22*(4), 545-585. https://doi.org/10.32890/jict2023.22.4.2

Deng, R., Benckendorff, P., & Gannaway, D. (2020). Learner engagement in MOOCs: Scale development and validation. *British Journal of Educational Technology*, *51*(1), 245-262. https://doi.org/10.1111/bjet.12810

Dragomir, V. D., & Dumitru, M. (2023). Two years into the COVID-19 pandemic: An analysis of learning outcomes and student engagement at an economics university. *Journal of Accounting Education*, *65*, 100871. https://doi.org/10.1016/j.jaccedu.2023.100871

Dubey, P., Pradhan, R. L., & Sahu, K. K. (2023). Underlying factors of student engagement to E-learning. *Journal of Research in Innovative Teaching & Learning*, *16*(1), 17-36. https://doi.org/10.1108/JRIT-09-2022-0058

ElSayary, A., Mohebi, L., & Meda, L. (2022). The impact of the relationship of social/emotional, cognitive, and behavioral engagements on developing preservice teachers' digital competencies. *Journal of Information Technology Education: Research*, *21*, 269-295. https://doi.org/10.28945/4982

Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. *Education and Information Technologies*, *29*, 22125-22145. https://doi.org/10.1007/s10639-024-12704-0

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

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59-109. https://doi.org/10.3102/00346543074001059

Ghosh, S. K., Janan, F., & Ahmad, I. (2022). Application of the classification algorithms on the prediction of student's academic performance. *Trends in Sciences*, *19*(14), 5070-5070. https://doi.org/10.48048/tis.2022.5070

Gkontzis, A. F., Kotsiantis, S., Tsoni, R., & Verykios, V. S. (2018). An effective LA approach to predict student achievement. *Proceedings of the 22nd Pan-Hellenic Conference on Informatics* (pp. 76-81). Association for Computing Machinery. https://doi.org/10.1145/3291533.3291551

Gledson, A., Apaolaza, A., Barthold, S., Günther, F., Yu, H., & Vigo, M. (2021). Characterising student engagement modes through low-level activity patterns. *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 88–97). Association for Computing Machinery. https://doi.org/10.1145/3450613.3456818

Guo, C., Yan, X., & Li, Y. (2020). Prediction of student attitude towards blended learning based on sentiment analysis. *Proceedings of the 9th International Conference on Educational and Information Technology* (pp. 228-233). Association for Computing Machinery. https://doi.org/10.1145/3383923.3383930

Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U., & Sattar, M. U. (2020). Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Applied Sciences*, *10*(11), 3894. https://doi.org/10.3390/app10113894

Hassan, H., Ahmad, N. B., & Sallehuddin, R. (2021). An empirical study to improve multiclass classification using hybrid ensemble approach for students' performance prediction. In R. Alfred, H. Iida, H. Haviluddin, & P. Anthony, P. (Eds.), *Computational Science and Technology* (pp. 551-561). Springer. https://doi.org/10.1007/978-981-33-4069-5_45

Helal, S., Li, J. Y., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J., & Long, Q. (2018). Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, *161*, 134-146. https://doi.org/10.1016/j.knosys.2018.07.042

Hollister, B., Nair, P., Hill-Lindsay, S., & Chukoskie, L. (2022). Engagement in online learning: Student attitudes and behavior during COVID-19. *Frontiers in Education*, *7*, 851019. https://doi.org/10.3389/feduc.2022.851019

Hsueh, N.-L., Daramsenge, B., & Lai, L.-C. (2022). Exploring the influence of students' modes of behavioral engagement in an online programming course using the partial least squares structural equation modeling approach. *Journal of Information Technology Education: Research*, *21*, 403-423. https://doi.org/10.28945/5010

Hussain, M., Hussain, S., Zhang, W., Zhu, W., Theodorou, P., & Abidi, S. M. R. (2018). Mining Moodle data to detect the inactive and low-performance students during the Moodle course. *Proceedings of the 2nd International Conference on Big Data Research* (pp. 133-140). Association for Computing Machinery. https://doi.org/10.1145/3291801.3291828

Ismail, N., & Yusof, U. K. (2022). Recent trends of machine learning predictions using open data: A systematic review. *Journal of Information and Communication Technology*, *21*(3), 337-381. https://doi.org/10.32890/jict2022.21.3.3

Ismail, S. N., Hamid, S., Ahmad, M., Alaboudi, A., & Jhanjhi, N. (2021). Exploring students engagement towards the learning management system (LMS) using learning analytics. *Computer Systems Science & Engineering*, *37*(1), 73-87. https://doi.org/10.32604/csse.2021.015261

Jamaluddin, A. H., & Mahat, N. I. (2021). Validation assessments on resampling method in imbalanced binary classification for linear discriminant analysis. *Journal of Information and Communication Technology*, *20*(1), 83-102. https://doi.org/10.32890/jict.20.1.2021.6358

Kabathova, J., & Drlik, M. (2021). Towards predicting student's dropout in university courses using different machine learning techniques. *Applied Sciences*, *11*(7), Article 3130. https://doi.org/10.3390/app11073130

Karalar, H., Kapucu, C., & Guruler, H. (2021). Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system. *International Journal of Educational Technology in Higher Education*, *18*(1), Article 63. https://doi.org/10.1186/s41239-021-00300-y

Khanal, S. S., Prasad, P., Alsadoon, A., & Maag, A. (2020). A systematic review: machine learning based recommendation systems for e-learning. *Education and Information Technologies*, *25*(4), 2635-2664. https://doi.org/10.1007/s10639-019-10063-9

Khatibsyarbini, M., Isa, M. A., Jawawi, D. N., & Tumeng, R. (2018). Test case prioritization approaches in regression testing: A systematic literature review. *Information and Software Technology*, *93*, 74-93. https://doi.org/10.1016/j.infsof.2017.08.014

Kitchenham, B. (2004). *Procedures for performing systematic reviews.* https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf

Kurniadi, F. I., Dewi, M. A., Murad, D. F., Rabiha, S. G., & Romli, A. (2023, November). An investigation into student performance prediction using regularized logistic regression. *Proceedings of the IEEE 9th International Conference on Computing, Engineering and Design, Kuala Lumpur, Malaysia*, 1-6. https://doi.org/10.1109/ICCED60214.2023.10425782

Li, X., Zhu, X., Zhu, X., Ji, Y., & Tang, X. (2020). Student academic performance prediction using deep multisource behavior sequential network. In H. Lauw, R. W. Wong, A. Ntoulas, E. P. Lim, S. K. Ng, & S. Pan (Eds.), *Advances in knowledge discovery and data mining* (pp. 567-579). Springer. https://doi.org/10.1007/978-3-030-47426-3_44

Lu, O. H. T., Huang, A. Y. Q., & Yang, S. J. H. (2021). Impact of teachers' grading policy on the identification of at-risk students in learning analytics. *Computers & Education*, *163*, 104109. https://doi.org/10.1016/j.compedu.2020.104109

Luo, J., & Wang, T. (2020). Analyzing students' behavior in blended learning environment for programming education. *Proceedings of the 2nd World Symposium on Software Engineering* (pp. 179-185). Association for Computing Machinery. https://doi.org/10.1145/3425329.3425346

Luo, N., Li, H., Zhao, L., Wu, Z., & Zhang, J. (2022). Promoting student engagement in online learning through harmonious classroom environment. *Asia-Pacific Education Researcher*, *31*, 541-551. https://doi.org/10.1007/s40299-021-00606-5

Ma, X., Yang, Y., & Zhou, Z. (2018). Using machine learning algorithm to predict student pass rates in online education. *Proceedings of the 3rd International Conference on Multimedia Systems and Signal Processing* (pp. 151-161). Association for Computing Machinery. https://doi.org/10.1145/3220162.3220188

Manzali, Y., Akhiat, Y., Abdoulaye Barry, K., Akachar, E., & El Far, M. (2024). Prediction of student performance using random forest combined with Naïve Bayes. *The Computer Journal*, *67*(8), 2677-2689. https://doi.org/10.1093/comjnl/bxae036

Marôco, J., Assunção, H., Harju-Luukkainen, H., Lin, S.-W., Sit, P.-S., Cheung, K.-c., Maloa, B., Ilic, I. S., Smith, T. J., & Campos, J. A. D. B. (2020). Predictors of academic efficacy and dropout intention in university students: Can engagement suppress burnout? *PLoS ONE*, *15*(10), e0239816. https://doi.org/10.1371/journal.pone.0239816

Muriana, I. O., Aiyegbusi, E., & Abam, S. (2022). 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

Nagarajan, H., Alsalami, Z., Dhareshwar, S., Sandhya, K., & Palanisamy, P. (2024, May). Predicting academic performance of students using modified decision tree based genetic algorithm. *Proceedings of the Second International Conference on Data Science and Information System, Hassan, India*, 1-5. https://doi.org/10.1109/ICDSIS61070.2024.10594426

Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, *11*(1), 237. https://doi.org/10.3390/app11010237

Nguyen, V. A., Nguyen, Q. B., & Nguyen, V. T. (2018). A model to forecast learning outcomes for students in blended learning courses based on learning analytics. *Proceedings of the 2nd International Conference on E-Society, E-Education and E-Technology* (pp. 35-41). Association for Computing Machinery. https://doi.org/10.1145/3268808.3268827

Orji, F., & Vassileva, J. (2020, September). Using machine learning to explore the relation between student engagement and student performance. *Proceedings of the 24th International Conference Information Visualisation, Melbourne, Australia*, 480-485. https://doi.org/10.1109/IV51561.2020.00083

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, *27*(6), 7893-7925. https://doi.org/10.1007/s10639-022-10925-9

Palanci, A., Yılmaz, R. M., & Turan, Z. (2024). Learning analytics in distance education: A systematic review study. *Education and Information Technologies*, *29*, 22629-22650. https://doi.org/10.1007/s10639-024-12737-5

Palli, A. S., Jaafar, J., Gilal, A. R., Alsughayyir, A., Gomes, H. M., Alshanqiti, A., & Omar, M. (2024). Online machine learning from non-stationary data streams in the presence of concept drift and class imbalance: A systematic review. *Journal of Information and Communication Technology*, *23*(1), 105-139. https://doi.org/10.32890/jict2024.23.1.5

Park, H. S., & Yoo, S. J. (2021). Early dropout prediction in online learning of university using machine learning. *International Journal on Informatics Visualization*, *5*(4), 347-353. https://doi.org/10.30630/joiv.5.4.732

Pooja, K. S. (2024). Machine learning advancements in education: An in-depth analysis and prospective directions. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(3), 3229–3237. https://www.ijisae.org/index.php/IJISAE/article/view/5928

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version*, *1*(1), b92.

Quigley, M., Bradley, A., Playfoot, D., & Harrad, R. (2022). Personality traits and stress perception as predictors of students' online engagement during the COVID-19 pandemic. *Personality and Individual Differences*, *194*, 111645. https://doi.org/10.1016/j.paid.2022.111645

Ramaswami, G. S., Susnjak, T., Mathrani, A., & Umer, R. (2020, December). Predicting students final academic performance using feature selection approaches. *Proceedings of the IEEE Asia-Pacific Conference on Computer Science and Data Engineering, Gold Coast, Australia*, 1-5. https://doi.org/10.1109/CSDE50874.2020.9411605

Ramon, Y., Matz, S. C., Farrokhnia, R., & Martens, D. (2021). *Explainable AI for psychological profiling from digital footprints: A case study of big five personality predictions from spending data.* PsyArXiv. https://doi.org/10.48550/arXiv.2111.06908

Rane, N., & Paramesha, M. (2024). Explainable Artificial Intelligence (XAI) as a foundation for trustworthy artificial intelligence. In D. Patil, N. L. Rane, P. Desai, & J. Rane (Eds.), *Trustworthy artificial intelligence in industry and society* (pp. 1-27). Deep Science Publishing. https://doi.org/10.70593/978-81-981367-4-9_1

Riestra-González, M., del Puerto Paule-Ruíz, M., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers & Education*, *163*, 104108. https://doi.org/10.1016/j.compedu.2020.104108

Rozi, A. F., Wibowo, A., & Warsito, B. (2023). Resampling technique for imbalanced class handling on educational dataset. *JUITA*, *11*(1), 77-85. https://doi.org/10.30595/juita.v11i1.15498

Sagala, T. N., Permai, S. D., Gunawan, A. A. S., Barus, R. O., & Meriko, C. (2022, December). Predicting computer science student's performance using logistic regression. *Proceedings of the 5th International Seminar on Research of Information Technology and Intelligent Systems, Yogyakarta, Indonesia*, 817-821. https://doi.org/10.1109/ISRITI56927.2022.10052968

Saleem, F., Ullah, Z., Fakieh, B., & Kateb, F. (2021). Intelligent decision support system for predicting student's e-learning performance using ensemble machine learning. *Mathematics*, *9*(17), 2078. https://doi.org/10.3390/math9172078

Salta, K., Paschalidou, K., Tsetseri, M., & Koulougliotis, D. (2021). Shift from a traditional to a distance learning environment during the COVID-19 pandemic: University students' engagement and interactions. *Science and Education*, 31, *93*-122. https://doi.org/10.1007/s11191-021-00234-x

Sheik Abdullah, A., Abirami, R. M., Gitwina, A., & Varthana, C. (2021). Assessment of academic performance with the e-mental health interventions in virtual learning environment using machine learning techniques: A hybrid approach. *Journal of Engineering Education Transformations*, *34*, 79-85. https://doi.org/10.16920/jeet/2021/v34i0/157109

Song, X., Li, J., Sun, S., Yin, H., Dawson, P., & Doss, R. R. M. (2021). SEPN: A sequential engagement based academic performance prediction model. *IEEE Intelligent Systems*, *36*(1), 46-53. https://doi.org/10.1109/MIS.2020.3006961

Stanimirovic, A. S., Nikolic, M. S., Jovic, J. J., Ristic, D. I. I., Corac, A. M., Stoimenov, L. V., & Peric, Z. H. (2024). Applying eXplainable AI techniques to interpret machine learning predictive models for the analysis of problematic internet use among adolescents. *Elektronika ir Elektrotechnika*, *30*(2), 63-72. https://doi.org/10.5755/j02.eie.36316

Subramainan, L., & Mahmoud, M. A. (2020). A systematic review on students' engagement in classroom: Indicators, challenges and computational techniques. *International Journal of Advanced Computer Science and Applications*, *11*(1), 105-115. https://doi.org/10.14569/IJACSA.2020.0110113

Tai, J., Ajjawi, R., Bearman, M., & Wiseman, P. (2020). Conceptualizations and measures of student engagement: A worked example of systematic review. In O. Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond,

& K. Buntins (Eds.), *Systematic reviews in educational research: Methodology, perspectives and application* (pp. 91-110). Springer. https://doi.org/10.1007/978-3-658-27602-7_6

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers and Education*, *143*, 103676. https://doi.org/10.1016/j.compedu.2019.103676

Torres, K. M., & Statti, A. (2023). What can data tell us? Using classroom data to determine student engagement. *International Journal of Curriculum Development and Learning Measurement*, *4*(1), 1-13. https://doi.org/10.4018/IJCDLM.320219

Turan, Z., & Karabey, S. C. (2023). The use of immersive technologies in distance education: A systematic review. *Education and Information Technologies*, *28*(12), 16041-16064. https://doi.org/10.1007/s10639-023-11849-8

Uliyan, D., Aljaloud, A. S., Alkhalil, A., Al Amer, H. S., Mohamed, M. A. E. A., & Alogali, A. F. M. (2021). Deep learning model to predict students retention using BLSTM and CRF. *IEEE Access*, *9*, 135550-135558. https://doi.org/10.1109/ACCESS.2021.3117117

Wang, Z., Cai, L., Chen, Y., Li, H., & Jia, H. (2021). The teaching design methods under educational psychology based on deep learning and artificial intelligence. *Frontiers in Psychology*, *12*, 711489. https://doi.org/10.3389/fpsyg.2021.711489

Wells, J., Spence, A., & McKenzie, S. (2021). Student participation in computing studies to understand engagement and grade outcome. *Journal of Information Technology Education: Research*, *20,* 385-403. https://doi.org/10.28945/4817

Xie, S.-t., Chen, Q., Liu, K.-h., Kong, Q.-z., & Cao, X.-j. (2021). Learning behavior analysis using clustering and evolutionary error correcting output code algorithms in small private online courses. *Scientific Programming*, *2021*(1), Article 9977977. https://doi.org/10.1155/2021/9977977

Yadav, B. R. (2024). The ethics of understanding: Exploring moral implications of explainable AI. *International Journal of Science and Research*, *13*(6), 1-7. https://doi.org/10.21275/SR24529122811

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*, Article 11. https://doi.org/10.1186/s40561-022-00192-z

Yamasari, Y., Nugroho, S. M. S., Yoshimoto, K., Takahashi, H., & Purnomo, M. H. (2020). Identifying dominant characteristics of students' cognitive domain on clustering-based classification. *International Journal of Intelligent Engineering and Systems*, *13*(1), 167-180. https://doi.org/10.22266/ijies2020.0229.16

Yañez, A. G. B., Alonso-Fernandez, C., & Fernandez-Manjon, B. (2023). Systematic literature review of digital resources to educate on gender equality. *Education and Information Technologies*, *28*(8), 10639-10664. https://doi.org/10.1007/s10639-022-11574-8

Yang, D., Wang, H., Metwally, A. H. S., & Huang, R. (2023). Student engagement during emergency remote teaching: A scoping review. *Smart Learning Environments*, *10*, Article 24. https://doi.org/10.1186/s40561-023-00240-2

Yang, Y., Hooshyar, D., Pedaste, M., Wang, M., Huang, Y. M., & Lim, H. (2020). Predicting course achievement of university students based on their procrastination behaviour on Moodle. *Soft Computing*, *24*, 18777-18793. https://doi.org/10.1007/s00500-020-05110-4

Zacharias, J., von Zahn, M., Chen, J., & Hinz, O. (2022). Designing a feature selection method based on explainable artificial intelligence. *Electronic Markets*, *32*(4), 2159-2184. https://doi.org/10.1007/s12525-022-00608-1

Zafar, S., & Nausheen, M. (2022). Influence of student engagement on academic challenges: Does engagement matter? *Global Educational Studies Review, VII*(II), 306-313. https://doi.org/10.31703/gesr.2022(VII-II).29

Zheng, Y., Gao, Z., Wang, Y., & Fu, Q. (2020). MOOC dropout prediction using FWTS-CNN model based on fused feature weighting and time series. *IEEE Access*, *8*, 225324-225335. https://doi.org/10.1109/ACCESS.2020.3045157

Zou, M. Y., Wang, T., Xu, H., Li, X. J., & Wu, X. (2020). Using process visualization and early warning based on learning analytics to enhance teaching and learning. In X. Sun, J. Wang, & E. Bertino (Eds.), *Artificial intelligence and security* (pp. 175-183). Springer. https://doi.org/10.1007/978-981-15-8083-3_16

AUTHORS

Chong Ke Ting is a PhD student and teaching assistant at the Faculty of Computing Universiti Teknologi Malaysia. She holds a Bachelor of Computer Science (Bioinformatics) with Honours from the Faculty of Computing, Universiti Teknologi Malaysia. Her research interests include student engagement, artificial intelligence, machine learning, feature selection, and big data.

Noraini Ibrahim is a senior lecturer in the Faculty of Computing, Universiti Teknologi Malaysia. She holds a PhD in Computer Science (Software Engineering) from Universiti Teknologi Malaysia. Her main research interests include software engineering (software evolution), requirement engineering, unified modeling language, software architecture, software development, and metamodeling.

Sharin Hazlin Huspi is a senior lecturer in the Faculty of Computing, Universiti Teknologi Malaysia. Her main research interests include natural language processing, machine learning, machine translation, text mining and analytics, data mining, and knowledge discovery.

Wan Mohd Nasir Wan Kadir is currently a professor and dean in the Faculty of Computing, Universiti Teknologi Malaysia. He has a PhD from the University of Manchester. He has been an academic staff member in the Faculty of Computing for over 27 years (since 1997). His main research interests include software engineering, software quality assurance, software testing, software architecture and design, requirement engineering, software verification and validation, software configuration management, artificial intelligence and machine learning in software engineering, and service-oriented computing.

Mohd Adham Isa is an associate professor in the Faculty of Computing, Universiti Teknologi Malaysia. He holds a PhD from the Universiti Teknologi Malaysia. His main research interests include software quality, software design, software analysis, unified modeling analysis, software testing, software architecture, software development, model-driven development, and quality assurance.