IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 December 2023, accepted 26 January 2024, date of publication 2 February 2024, date of current version 16 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3361479

SURVEY

Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review

LIDYA R. PELIMA, YUDA SUKMANA, AND YUSEP ROSMANSYAH
School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung 40132, Indonesia

Corresponding author: Lidya R. Pelima (23221104@mahasiswa.itb.ac.id)

This work was supported in part by the (Lembaga Pengelola Dana Pendidikan (LPDP)/Indonesia Endowment Fund for Education Agency) Riset Inovatif Produktif (RISPRO) Invitasi, and in part by the Statistics Indonesia-Badan Pusat Statistik (BPS).

ABSTRACT Predicting university student graduation is a beneficial tool for both students and institutions. With the help of this predictive capacity, students may make well-informed decisions about their academic and career paths, and institutions can proactively identify students who may not graduate and offer tailored support to ensure their success. The use of machine learning for predicting university student graduation has drawn more attention in recent years. Large datasets of student academic performance data can be used to train machine learning algorithms to identify patterns that are applicable in predicting future outcomes. In accordance with some studies, this approach predicts student graduation with an accuracy rate as high as 90%. Many systematic literature reviews (SLRs) have been conducted in this field, but there are still limitations, including not discussing the predictive models and algorithms used, a lack of coverage of the machine learning algorithms applied, small database coverage, keyword selection that does not cover all synonyms relevant to the investigation, and less specific data collection transparency. By delving into the limitations of existing SLRs on this topic, this research not only enhances the understanding of machine learning applications in forecasting student graduation but also fills a crucial gap in the literature. The inclusion of weaknesses in current SLRs provides a foundation for justifying the need for this study, emphasizing the necessity of a more nuanced and comprehensive review to advance the field and guide future research efforts in smart learning environments. This research conducts a thorough systematic review of the existing literature on machine learning-based student graduation prediction models from 70 journal articles from 2018 through 2023 that are pertinent. This review includes the various machine learning algorithms that have been implemented, the various academic performance data that was obtained from students, and the effectiveness of the models that have been developed. It also discusses the difficulties and potential advantages of utilizing machine learning to predict student graduation. The review indicates that the most common approach employed is the prediction of students' academic performance, which relies on data obtained from the Learning Management System and Student Information System. The primary data utilized for prediction purposes consists Student retention and time of academic and behavioral information. Among the various algorithms employed, Support Vector Machine and Random Forest are the most commonly utilized. This study makes a significant contribution to the advancement of learner modules within the smart learning environment.

INDEX TERMS SLR, academic performance prediction, higher education, machine learning.

The associate editor coordinating the review of this manuscript and approving it for publication was Laxmisha Rai.

I. INTRODUCTION

University graduation is a significant life achievement, but it is not always simple. Student retention and the time to

graduation continue to cause concerns for administrators and faculty members. Longer graduation times put more financial strain on students and the university's limited assets [1]. Graduation on time means finishing all the necessary classes and earning the required number of credits to graduate within the number of years that is typically expected for a particular degree program [2].

A crucial part of developing new leadership is training students from higher education institutions (HEIs). It equips them with the tools they need to challenge accepted paradigms, foster new thinking in line with contemporary issues and traits, achieve continuing, self-directed learning, and adjust to a variety of work-related circumstances [3]. The importance of student graduation is based on the aforementioned factors.

Although there have been many SLRs conducted in this field, there are still some research gaps such as studies that do not discuss in depth the predictive models and algorithms used in machine learning [4], lack of exploration of what types of data are used to improve the accuracy and comprehensiveness of predictive models, lack of discussion of various machine learning algorithms and their application in predicting student success [5], SLRs that only focus on purely online courses, which may limit the generalisation of findings to other types of courses or learning environments [6], less comprehensive coverage of source databases [7], [8], keyword selection which did not include all possible variations or synonyms relevant to the investigation and lack of transparency in data collection [8].

The emergence of educational database management systems has resulted in the creation of numerous educational databases facilitating data mining to extract valuable insights from this data [5]. The COVID-19 pandemic has significantly contributed to the adoption of technology in the field of education. In an online statistics course, pass rates and final test results are predicted using data analysis. Utilizing previously collected data, the goal is to pinpoint students who are more likely to fail their final examinations or dropout [9]. In an online learning environment, the Hidden Markov model is also applied to examine and model sequential student learning patterns. The main objectives are to identify at-risk students early and provide necessary interventions [10]. Homework assignment collection data is used to predict students' academic tendencies to postpone assignments during blended learning courses. The purpose is to identify students who procrastinate and establish appropriate interventions methods to help them [11]. Machine learning techniques are used to forecast student academic achievement in a smart campus setting, specifically to discover the characteristics that contribute to student success and execute suitable strategies to improve their achievement [12]. The early identification of students with the possibility of failing in face-to-face courses is also applicable [13]. To boost the prediction performance of students who are likely to drop out, a two-layer ensemble machine learning method is employed [14]. Investigation by [15] employed an ensemble machine learning technique termed stacking to identify early learners who were at risk of dropping out. By combining several prediction models, this technique improves the accuracy as well as reliability of its forecasts. In order to identify students who may struggle, [16] uses a multitask learning strategy based on multi-instance multi-label learning to predict student performance prior to the start of the course. An empirical study that aims to predict the academic performance of Master's program students in Germany incorporates various factors, which include demographic data, post-enrollment attributes such as grades, the number of failed courses, the number of registered and unregistered exams, the distance from students' accommodation to the university, cultural data, and other supporting data to build a predictive model [17]. Analyzing patterns of behavior that occur on campus may also be applied to predict academic success. The concept is to find and exploit behavior patterns associated with academic success to predict student performance [18]. Early warning systems are being incorporated as a predictor of student performance in blended learning courses in higher education. The target is to create a system that is able to detect low-risk students at the beginning of the course [19]. To construct an effective predictive model for determining students who face the possibility of dropping out of university, innovative statistical and multilevel machine learning can be applied for forecasting as promptly as feasible [20]. The application of educational data mining (EDM) techniques can be used to measure how well the various algorithms predict student graduation rates in higher education institutions [21]. Examining the impact of demography on student success with the use of early-warning systems to intervene with low-risk students can help identify the demographic aspects that affect students' performance. The next phase is to establish effective intervention options [22]. Gender, prior achievement, attendance at lectures, engagement in tutorials, and performance on tests are utilized to investigate patterns associated with students' success or failure as first-year university students [23].

Academic performance is the degree to which a student has met his or her short- or long-term educational objectives [18]. It has been consistently proven that the following elements can have a major effect on academic performance: personality of students, personal status, lifestyle behaviors, learning behaviors [24], high school preparation, and socioeconomic background [25].

EDM is a new discipline focused on developing techniques for studying the distinctive and progressively massive amounts of data generated by educational institutions and applying these techniques to gain a better understanding of students and the environments in which they learn. Regardless of whether the educational data is derived from students' use of interactive educational environments, computer-supported collaborative learning, or administrative information from schools or universities, it tends to have multiple levels of meaningful hierarchy, which must often be identified through data properties rather than beforehand.

Time, chronology, and context are other essential considerations in the study of educational data [26], [27].

EDM has sparked the interest of scholars, encouraging the completion of numerous systematic literature studies, among others including research to identify major trends, study themes, and influential writers in the area of EDM, investigate the effectiveness and weaknesses [26], [28], conduct a detailed evaluation of machine learning methodologies used for predicting student dropout in an online course by analyzing various algorithms and procedures [29], and a study that investigates the various uses of machine learning in the educational sector and examines the influence on student learning results [30].

To gain a deep and thorough understanding of how academic performance can be used in student assessment, research was conducted in the form of a systematic literature review (SLR). An SLR is a method of identifying, evaluating, and summarizing the existing research literature on a particular topic. This literature review uses the results of research in the field of EDM from the last five years (2018-2023). The review includes the data sources used in the research, the methods, the variables selected for the prediction stage, and the software used. The SLR is divided into five sections, the first of which is about the preparation of the SLR. The second section will go through the research technique in depth, while the third section will go over the findings and responses to the study questions. The fourth section will analyze the limitations and obstacles faced, and the last part will summarize the research findings and provide recommendations for future research based on the findings. The present study provides a substantial contribution to the progress of the learner module in smart learning environments.

II. RESEARCH METHOD

A systematic literature review methodology based on Kitchenham's recommendations is used in this research. The recommendations consist of three stages, as follows:

Planning Stage: It is essential to confirm the need for a systematic literature review before initiating it. At this stage, the goals are to (a) recognize the necessity for a review, (b) initiate a review, (c) define the research question(s), and (d) establish the review protocol.

Conducting Stage: There are five steps taken in this stage, namely:

A. Identification of research. This step is to discover as many relevant research articles and studies as possible that are related to the topic.
B. Selection of primary studies. After the relevant research papers have been discovered, the next step is to choose the main articles that meet the inclusion criteria.
C. Study quality assessment. To ensure that the review contains high-quality studies, it is necessary to evaluate the chosen research papers.
D. Data extraction and monitoring. The next step is to extract the relevant data from the chosen research paper. The extracted data is then organized and monitored carefully for further analysis.
E. Data Synthesis. Analyzing and summarizing the findings to draw meaningful conclusions and identify any patterns or trends is the last step in this research stage.

Reporting Stage: This last stage aims to communicate the findings and results of the review effectively to related parties.

A. RESEARCH QUESTIONS

Research questions are the most significant aspect for a reviewer to identify and address in SLR [31]. Throughout the review, we attempt to recognize and solve the research questions based on Kitchenham's point of reference [32]. PICOC is an acronym for Population, Intervention, Comparison, Outcome, and Context. An explanation of the PICOC for this research is as follows:

The target population for the data collection is specifically referring to the group of individuals, programs, or businesses that are the focus of the review. The population under investigation in this study comprises the fields of educational data mining, educational data analytics, and learning analytics.

The term ''intervention'' pertains to the specific methodology, tool, technology, or procedure within the field of software engineering that is currently under evaluation. The study incorporates machine learning as an intervention.

The term ''comparison'' refers to the manner in which the intervention and control conditions are distinguished within the studies incorporated in the review. This review encompasses an analysis of student performance and graduation rates as the basis for comparison.

The outcome corresponds to the quantification of the impact resulting from the implementation of the intervention. The study's findings cover the prediction of student dropouts, the development of an early warning system for identifying potential dropouts, and the prediction of student graduation.

The term ''context'' applies to the specific environment or circumstances in which the intervention under investigation was examined. The contextual elements encompassed in this study may comprise various factors, including but not limited to the organizational type, organizational size, and development process, as well as the tools and techniques employed. This review is situated within the scope of higher education.

Based on the PICOC, the following research questions were obtained:

- RQ1: What is the state-of-the-art research conducted in educational data mining (EDM) and educational data analytic (EDA) in higher education?
- RQ2: What are the sources of data, the variables, and the collection techniques that have been used to predict students' academic performance in higher education using EDM?
- RQ3: What are the gaps and challenges of research conducted in EDM?

B. SEARCH STRATEGY

Prior to commencing the study, it is imperative to carefully choose and substantiate a search strategy for solving the research questions. This entails a cautious evaluation of the keywords and terminology employed in the search for academic literature, as well as the databases and other resources that will be explored. The formulation of the search strategy is customized to suit the particular research query. After formulating the search strategy, it is required to conduct a comprehensive search across a wide range of important electronic sources. Academic databases, such as ACM digital library, IEEE digital library, Science Direct, Springer Link and Taylor and Francis are encompassed within this study. By complying with the prescribed methods, we can enhance the likelihood of locating a maximum number of primary studies that are relevant for solving the research questions.

Once the database has been identified, the subsequent step involves compiling a comprehensive inventory of synonyms, abbreviations, and alternative spellings derived from the PICOC framework, research questions, as well as significant relevant terms found within titles and abstracts. This process is undertaken to facilitate the selection of appropriate keywords. The study employs the following keywords: dropout prediction, early warning system of student dropout, educational data analytics, educational data mining, learning analytics, machine learning, student graduation, student graduation prediction, student performance, higher education. Advanced keyword searches can subsequently be formulated by utilizing boolean operators such as AND and OR. The search terms that were generated and utilized in this literature review are (‘‘educational data analytics’’ OR ‘‘educational data mining’’ OR ‘‘learning analytics’’) AND (‘‘machine learning’’) AND (‘‘student graduation’’ OR ‘‘student performance’’) AND (‘‘dropout prediction’’ OR ‘‘early warning system of student dropout’’ OR ‘‘student graduation prediction’’) AND (‘‘higher education’’).

C. STUDY SELECTION CRITERIA

The papers that are included in the review are chosen based on certain criteria. These criteria are settled prior to the start of the review and are based on the review’s research question. The Parsif.al and Zotero applications serve for managing and storing search result articles. For this review, the criteria are:

- Inclusion: papers written in English; acceptable study types include empirical studies, practical studies, and mixed studies. Journal articles published between 2018 and 2023 that fall under the Q1-Q3 publication rankings range. Concerning the subjects of educational data analytics, educational data mining, and learning analytics. Concentrated on student performance prediction or student graduation prediction. Target students in higher education and utilize machine learning.
- Exclusion: papers not written in English, conference papers, books, systematic literature reviews, and other types of publications published prior to 2018, not related to the topic of educational data analytics, educational data mining, learning analytics, not related to student performance prediction or student graduation prediction, not related to the scope of a lecturer in higher education, and mainly utilizing deep learning.

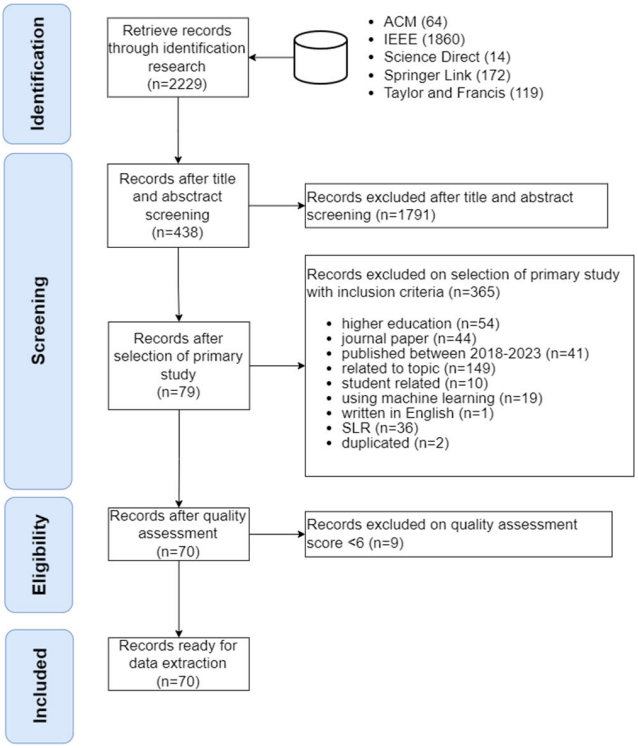

FIGURE 1. Flowchart for search and selection methodology.

D. QUALITY ASSESSMENT

A key element in any systematic literature review (SLR) is quality assessment. It involves evaluating the primary research that is incorporated into the review’s strategy. The goal of quality assessment is to guarantee the accuracy and dependability of the SLR’s conclusions. It can be conducted by delivering a list of questions in the form of a survey. This is in accordance with the [32] recommendations:

- Is a clear description of the findings present?
- Does the publication provide an assessment of the results?
- Are the limits discussed in the paper?
- Does the study mention upcoming projects?
- Does the study have any practical or research value?
- Is the research’s history and objectives clearly stated?
- Does the research methodology have a description?
- Are the data source and collection described in the paper?
- Is a well-defined variable being used?

TABLE 1. Quality assessment checklist.

Category	Description	Score
Yes	The content is presented in a clear and explicit manner	1
Partially	The content is presented in a clear but implicit manner	0.5
No	The content is inconclusive	0

The answers to the evaluation of the above questions are broken down into three groups as shown in Table 1. The question will receive a score of 1 if the affirmative response states that there is a thorough and understandable explanation. A question will receive a score of 0.5 if it receives a partial or not complete and clear reply; if it receives no response, it will receive a weight of 0. Additionally, the publication's importance will be determined by the response to the query. The maximum possible score is 9, while the minimum possible score is 0. The publications that will be attentively and thoroughly studied will then be decided upon by this final evaluation. Based on the results of the assessment, the threshold value for publications to be used in this study is 6. A fair and thorough assessment of the reviewed content is required, which is why a cut-off score of 6 with a maximum possible score of 9 and 9 total questions were chosen for the quality assessment stage. This cutoff point guarantees a rigorous procedure by requiring material to fulfill a substantial amount of the requirements, demonstrating a dedication to superior standards. In a balanced scoring system, a score greater than 6 denotes a two-thirds threshold and suggests a solid understanding of the evaluation criteria. A score of six or higher indicates thorough comprehension and efficient implementation of the criteria, permitting only excellent reviews to add to the final evaluation. This criterion preserves assessment by preventing the inclusion of reviews that do not meet fundamental features while upholding rigor and reliability. Out of a total of 79 publications, 9 publications were omitted because they did not meet the rating threshold, so only 70 publications would be read carefully and thoroughly.

TABLE 2. Publication category.

No	Publication Category	Qty
1	Q1	66
2	Q2	3
3	Q3	1

Table 2 contains the comprehensive collection of primary research publications that were ultimately selected for inclusion. These articles have been organized and classified based on the respective Q category descriptors, which are denoted as Q1, Q2, and Q3. Out of the 70 publications that were selected, 66 belong to the Q1 category, 3 fall under the Q2 category, and 1 belongs to the Q3 category. There are no publications that have been specifically chosen or designated for the Q4 category.

E. DATA EXTRACTION

During the data extraction phase, valuable information is documented in the chosen journal based on the predetermined description established during the planning phase. The aforementioned data is subsequently utilized to address inquiries within the context of research. The data is documented in a tabular format, which enables subsequent analysis. In this data extraction stage, a total of 15 variables are employed. These variables consist of the publication title, year of publication, journal source database, journal title, journal address, Scopus index, research objective, research data source, method or algorithm employed, variable or type of selected data, tools utilized, discussion and results, weaknesses identified, and suggestions for future development.

III. RESULT AND DISCUSSION

This section provides insight into the findings and offers a comprehensive discussion. It begins with an overview of how the research paper was distributed, followed by a discussion of the findings for each research question.

A. DESCRIPTION OF STUDIES

This review incorporates scholarly journals that have been published between the years 2018 and March 2023. A total of 70 journals were obtained from the previous stages. The distribution of these journals by year of publication is shown in Figure 1 as follows: 8 selected journals were published in 2018, 15 selected journals were published in 2019, 12 selected journals were published in 2020, 13 selected journals were published in 2021, 19 selected journals were published in 2022, and as of May 2023, there were 3 selected journals available. This finding indicates that there continues to be a significant level of interest among scholars in this academic field. The year 2020 witnessed a decline in the quantity of chosen publications, a trend that can be attributed to the impact of the COVID-19 pandemic. The ongoing global pandemic has served as a catalyst for advancements in the domain of educational data mining and educational data analytics, leading to a notable surge in the adoption of learning management systems (LMS).

At the starting point of the study selection phase, a total of 438 articles have been identified for subsequent scrutiny, which include the following details: 53 papers were obtained from ACM, 225 publications were retrieved from IEEE, 14 publications were accessed from Science Direct, 31 publications were acquired from Springer Link, and 115 publications were obtained from Taylor and Francis. Following the completion of the study selection stage, which involved the exclusion of duplicate publications and those that did not satisfy the set criteria, the subsequent

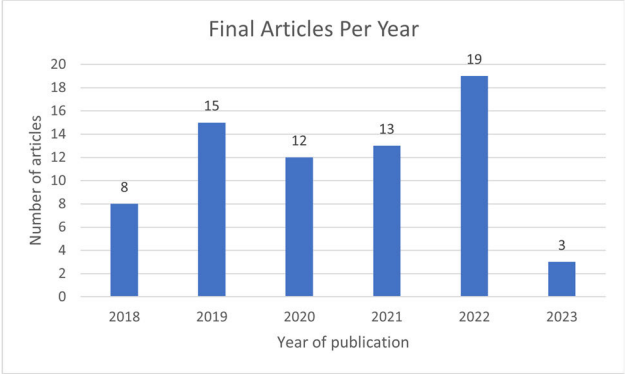

FIGURE 2. Final articles per year.

phase involved the quality assessment of the selected publications. The ultimate outcomes of this phase are shown in Table 3. A total of 2 publications were obtained from ACM, 43 publications from IEEE, 2 publications from Science Direct, 15 publications from Springer Link, and 8 publications from Taylor and Francis.

TABLE 3. Number of primary studies based on sources.

No	Digital Library	Qty	Papers
1	ACM	2	[1], [18]
2	IEEE	43	[9], [10], [11], [12], [13], [24], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70]
3	Science Direct	2	[14], [71]
4	Springer Link	15	[15], [17], [72], [73], [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84]
5	Taylor and Francis	8	[19], [20], [21], [22], [23], [85], [86], [87]

In the domain of databases, it is evident that IEEE holds the dominant position as the primary contributor to the entirety of the journal, accounting for 60% of its content. Springer Link holds the second-highest position as a contributor, accounting for 21% of the total. Taylor and Francis accounted for 11% of the overall number of journals, while both ACM and Science Direct individually contributed 3% of the total journals.

B. RQ1: WHAT IS THE STATE-OF-THE-ART RESEARCH CONDUCTED IN THE AREA OF EDM AND EDUCATIONAL DATA ANALYTIC IN HIGHER EDUCATION

The initial study inquiry examines existing research patterns in the domains of educational data mining and educational data analysis, with a particular focus on investigations done within the environment of higher education. The selected primary studies can be categorized into four distinct approaches based on their objectives. These approaches include Academic Performance Prediction, At-risk Student Prediction, Dropout/Graduation Prediction, and Learning Analysis with its application in the academic domain. Based on the data presented in Table 4, it is evident that 23 scholarly articles focus on the prediction of academic performance, while 16 publications center around the prediction of at-risk students. Additionally, 14 publications delve into the topic of dropout and graduation prediction, and 17 publications explore the field of learning engagement.

TABLE 4. Mapping of approaches.

No	Approach	Paper
1	Academic Performance Prediction	[11], [12], [17], [18], [23], [24], [35], [37], [38], [41], [44], [45], [51], [56], [57], [64], [71], [74], [75], [78], [80], [81], [84]
2	At-Risk Student Prediction	[1], [9], [10], [13], [19], [36], [50], [55], [59], [62], [65], [66], [73], [77], [79], [82]
3	Dropout/Graduation Prediction	[14], [15], [20], [21], [47], [48], [54], [60], [61], [63], [72], [83], [85], [86]
4	Learning Engagement	[22], [34], [39], [40], [42], [43], [46], [49], [52], [53], [58], [67], [68], [69], [70], [76], [87]

In accordance with the findings presented in [23], academic performance pertains to the achievement of students in their pursuit of higher education, typically assessed through their academic grades or performance on examinations and assessments [18], [23], [34], [44], [63], [79], with particular emphasis on their initial year of study. Additionally, it encompasses the probability of students' persistence in pursuing further education or their decision to discontinue their academic pursuits. Several factors have been identified as influential in determining academic performance. These factors encompass prior academic achievement [23], [40], [43], demographics [23], [24], [36], psychological variables such as perceived confidence [23], [24], [37], and early engagement with the course and broader university environment [31], [32].

The prompt identification of students' academic performance is widely recognized as a vital asset for enhancing

the quality of education [70]. Researchers are showing interest in predicting academic performance, this can be seen from increasing numbers of research on predicting academic performance as well as selected primary publications stating that it is beneficial to students [23], [24], [34], [43], [44], [50], [70], [73], [77], [79], [83], instructors [11], [12], [17], [18], [23], [24], [36], [37], [40], [43], [44], [56], [63], [74], [77], educational institutions [17], [23], [24], [36], [77].

From the discussion above, it can be concluded that there are several objectives behind predicting academic performance:

- Early Interventions. The early identification of students who are at-risk of academic underperformance or dropping out enables educators and educational institutions to promptly intervene and provide the necessary support. The implementation of timely interventions has the potential to provide students with the essential resources and assistance required to enhance their academic achievements [11], [17], [18], [23], [36], [37], [43], [50], [55], [56], [63], [70], [73], [77], [79], [80], [83].
- Resource Allocations. Educational institutions frequently encounter constraints on their available resources. Through the utilization of academic performance prediction, resources can be allocated in a more efficient manner, enabling their directed allocation towards students or regions that require them the most [17], [74].
- Personalized Learning Experiences. The utilization of predictive insights has the potential to enhance the creation of customized or adaptable learning experiences that suit the unique needs of individual students. For instance, students who are anticipated to encounter difficulties in a specific academic discipline should be offered further materials or alternate pedagogical approaches [11], [18], [24], [40], [44], [50].
- Data-driven Decision Making. Predictive models offer instructors, administrators, and officials with data-driven information that can enhance decision-making in areas such as the development of curriculum, teaching practices, and institutional regulations [12], [18], [23], [74], [77].
- Enhance Educational Research. The prediction of academic success yields vital data for the field of educational research. This study aims to enhance comprehension of the determinants that impact student achievement and the intricate interplay between many variables [23], [44], [74].
- Economic Implications. The occurrence of students terminating their education or displaying poor academic achievement carries economic consequences, impacting both educational institutions in terms of funding and resource allocation as well as society due to the potential decrease in lifetime wages and lessened economic productivity for these students [17].
- Student Welfare and Mental Health. There is a strong correlation between academic difficulties and mental health issues. By proactively identifying and tackling academic obstacles in their early stages, educational institutions have the ability to minimize subsequent mental health concerns [11], [18].
- Technological Advancements. The rapid growth of educational technology, incorporating LMS and online education platforms, has resulted in the availability of a significant amount of student-centric data. The extensive dataset in this context offers researchers the chance to employ sophisticated algorithms and extract significant predictions [34], [44], [74], [79].
- Competitive Edge for Institutions. In the context of the highly competitive educational landscape, the ownership of sophisticated predictive systems may bring a distinct advantage to educational institutions. These systems enable institutions to anticipate and plan for many factors, such as funding, student enrollment, and reputation, thereby enhancing their competitive position. The institution's dedication to student achievement and contemporary, evidence-based methodology is exemplified by this display [34], [77].
- Long-Term Success Metrics. In addition to short-term academic achievements, predictive models have the capacity to anticipate long-term indicators of success, such as professional advancements, an inclination for lifelong learning, and various outcomes beyond the completion of education [17].

Currently, the prediction of student achievement has emerged as a prominent area of research. This is mostly due to its significant influence on enhancing students' academic performance. Various educational data mining approaches are being employed to offer essential assistance to students who are at-risk in their studies [49]. At-risk students are one of the main focuses of universities and lecturers [1]. At-risk students, based on various publications, can be defined as individuals who may be prone to discontinuing their enrollment or facing unsuccessful outcomes in their current course or program [9], [10]. Academic achievement for these students relies upon the assistance and involvement of instructors and educational systems [19], [49], [81]. They may exhibit distinct behavioral patterns, including a tendency to solely observe a course without actively participating, consistently interacting but with poor outcomes, or displaying irregular participation that may result in high-risk situations [10], [64]. Recognition of at-risk students involves considering various parameters, encompassing socio-demographic characteristics, academic achievement, behavioral patterns, previous performance, and particular criteria pertaining to subgroups, such as race or socioeconomic background [58]. Classification of at-risk students is often based on their academic performance, where final grades are converted into binary indicators of success or being at risk in accordance with established academic criteria [1], [65], [76]. The early identification of at-risk students has advantages for educators, learners, and academic institutions [1]. Educators acquire knowledge

regarding the specific students that require further support in learning, thereby enabling the timely provision of aid to enhance their academic performance [13], [35], [54], [58], [65], [72], [76]. Students are able to improve their academic performance [1], [9], [10], [49], [54], [58], [64], [72], [76], [78], [81]. Academic institutions experience advantageous outcomes such as diminished rates of dropping out, which in turn enables them to obtain timely information for effective decision-making [10], [35], [54], [72], [81].

Within the subject of higher education, there are two distinct categories of dropout phenomena: (a) course-level dropout, and (b) beyond course-level dropout [9]. In the aforementioned scenario, the act of discontinuing participation in a course or subject takes place, providing an opportunity for educators to intervene and mitigate dropout rates by leveraging pertinent information. In addition to dropouts occurring at the course level, students may also choose to withdraw from their academic studies. The term ''dropout'' in the context of higher education refers to students who discontinue studies prior to the successful completion of their degree program [15], [53]. The student does not engage in the process of transferring to another university or enrolling at the university during the subsequent year [60]. The voluntary withdrawal is typically motivated by factors such as the desire to change academic majors or transfer to different educational institutions [14], [20]. A student is classified as a dropout in an online course when they cease to engage in any learning activities within a designated timeframe. Academic indicators of non-compliance may appear through the omission of assignment submissions, a lack of observable learning behaviors within a designated timeframe, or the absence of log data [14], [59], [82]. Failing to acquire a course certificate is also considered a dropout [47], [62]. Graduation can be defined as the attainment of a degree within the prescribed timeframe [21], [53], [85] by a student enrolled at a university [20], whether it corresponds to their original major they were registered for or not [84], and is typically marked by a ceremony [71].

Educational analytics also fulfills the function of evaluating students' learning behavior and degree of involvement. This assessment facilitates the implementation of essential alterations to the curriculum, enhancements to its content, and adaptations to the teaching technique, all with the objective of promoting students' academic achievement. Within the realm of this particular academic discipline, scholarly literature predominantly centers its attention on three fundamental behaviors, namely involvement, committed efforts, and participation [87]. The discipline of learning analytics encompasses various dimensions and focuses on the methodical gathering, examination, and presentation of data regarding learners and their educational context. The principal objective is to comprehend and enhance both the process of acquiring knowledge and the environment in which it occurs [33], [38], [39], [42], [51], [57], [68], using advanced artificial intelligence techniques [67], [75].

C. RQ2: WHAT ARE THE SOURCES OF DATA, THE VARIABLES, AND THE COLLECTION TECHNIQUES THAT HAVE BEEN USED TO PREDICT STUDENTS' ACADEMIC PERFORMANCE IN HIGHER EDUCATION USING EDM

The majority of research associated with the prediction of academic achievement has been conducted with data that was collected at the completion of the academic semester and derived from multiple sources [78]. Table 5 displays the data sources utilized in the chosen publication.

TABLE 5. Data sources.

No	Data Source	Paper
1	LMS	[1], [9], [11], [12], [13], [14], [19], [35], [36], [37], [38], [39], [40], [43], [44], [45], [54], [58], [63], [65], [66], [68], [69], [70], [71], [73], [74], [77], [79], [84], [87]
2	SIS	[15], [17], [21], [22], [54], [55], [56], [62], [76], [78], [85], [86]
3	Hybrid/LMS and SIS	[10], [18], [20], [23], [24], [34], [46], [47], [48], [49], [50], [51], [52], [53], [57], [59], [60], [61], [72], [72], [75], [81], [82]
4	Other	[41], [42], [64], [80], [83]

Hybrid data sources are classified as such when they encompass a combination of data obtained from both LMS and Student Information System (SIS) sources. This study categorizes public data sources such as the Open University Learning Analytics Dataset (OULAD) and KDDCUP as hybrid due to their inclusion of demographic information, such as age and gender, obtained from the SIS, as well as interaction and grade data obtained from the LMS [10], [47], [59]. The data extracted from the LMS consists of learning activity data, encompassing many aspects such as course participation, login frequency, time spent reading or viewing lecture materials, engagement in discussion forums, homework, performance in quizzes, mid-term tests, and end-of-term exams [51]. The data obtained from the SIS mostly consists of personal information pertaining to students. However, this data is still relevant to the educational context as it includes demographic information, admission records and prior education records, as well as data on courses completed in the previous semester, including corresponding grades and Grade Point Average (GPA) [17]. Additional data sources commonly utilized in research include the implementation of questionnaires to collect responses for subsequent analysis, as employed by [40] and [41], as well as the utilization of social media data, as employed by [63].

On the basis of the findings of a comprehensive study that involved the analysis of particular research publications,

it has been noticed that a wide variety of data is exploited in order to forecast student performance. Table 6 provides a concise summary of these findings.

TABLE 6. Student-level predictor data type.

No	**Student-Level Predictor** Data Type	Paper
1	Academic	[9], [13], [34], [35], [37], [38], [39], [40], [42], [44], [45], [54], [62], [72], [76], [78], [86]
2	Behavioral	[11], [19], [24], [43], [58], [60], [63], [64], [65], [69], [71], [73]
3	Academic and Behavioral	[1], [10], [12], [14], [18], [46], [47], [48], [49], [51], [53], [57], [68], [74], [77], [79], [80], [84]
4	Academic and Demographic	[17], [36], [56]
5	Academic, Demographic and Pre-University	[20], [21], [22], [41], [50], [82], [85]
6	Academic, Demography and Behavioral	[52], [59], [61], [66], [67], [75], [81], [83], [87]
7	Academic, Pre-university and Admission Test	[55]
8	Academic, Demographic, Pre-university, and Admission Test	[15]
9	Academic, Demographic, Pre-university, and Behavior	[23]

Academic data represents the data that is produced throughout the educational journey of students [39]. This data comprises many attributes that depict students' performance in a particular course, such as class information, assignment details, grades, and attendance records [9], [34], [36], [37], [43], [43], [53], [71], [75], [77], [85]. These attributes tend to be displayed in the form of tabular and time-series data [61]. Behavioral data extends to the collection of information concerning students' activities and behaviors, which are obtained from both LMS and physical interactions [24]. This data includes various aspects such as frequency of learning, number of clicks, logins, time spent, emotional responses during online learning, interactions within the library, Wi-Fi data indicating students movements, calculations related to attendance [11], [19], [57], [59], [63], [64], [68], [70], and is typically presented in the format of log files [19], [62], [70]. Demographic data incorporates several socio-demographic factors, including but not limited to gender, age, ethnicity, major of study, marital status, household income, high school GPA, full-time or part-time student status, education level of parents, and financial reliance status. The financial reliance status refers to the extent to which a student relies on their parents' financial assistance [13], [18], [44], [45], [49], [57], [63], [71], [79]. These characteristics are classified as static as they do not require frequent updates or revisions [65].

Pre-university data represents a range of information regarding students' attributes and achievements prior to their enrollment in the context of a university. This incorporates data on their educational trajectory, personal and demographic details [20], academic performance in secondary school and higher secondary school, pre-college and pre-program participation, scores on the National Achievement Test (NAT), admission test scores, marks obtained in intermediate and matriculation programs [15], [20], [22], [23], [40], [49], social interaction network, and programming knowledge [54].

In order to predict student performance, multiple algorithms are employed.

TABLE 7. Learning algorithms.

No	Learning Algorithm	Paper
1	Support Vector Machine	[1], [8], [9], [13], [20], [32], [34], [37], [38], [40], [45], [46], [48]–[53], [56], [59], [62], [66]–[69], [72]–[75], [77]
2	Random Forest	[11], [14], [17], [18], [20], [23], [24], [38], [44], [47], [48], [50], [51], [52], [56], [57], [60], [63], [64], [66], [67], [68], [69], [77], [78], [79], [81]
3	Artificial Neural Network/ Multi Layer Perceptron	[12], [13], [17], [18], [41], [50], [55], [56], [62], [65], [71], [72], [73], [74]
4	Logistic Regression	[12], [13], [15], [17], [38], [56], [63], [68], [71], [73], [81], [85], [87]
5	K-Nearest Neighbors	[12], [13], [36], [38], [42], [44], [52], [56], [78]

In addition to examining the nature of data, this review also explores the prevailing and often employed learning algorithm. Upon deeper examination, it is possible for a single research publication to go into many learning algorithms. According to the findings presented in Table 7, the analysis reveals the Support Vector Machine (SVM) was the most frequently employed, appearing in 31 publications. Following SVM, Random Forest algorithm was utilized in 27 papers, followed by the Artificial Neural Network/Multi-Layer Perceptron. Additionally, Logistic Regression was employed in 13 papers, and k-Nearest Neighbors (k-NN) was utilized in 9 publications.

D. RQ3: WHAT ARE THE GAPS AND CHALLENGES OF RESEARCH CONDUCTED IN EDM

It is evident from the review that the prediction of academic performance in higher education holds promise for assisting students, educators, and academic institutions. However, the

conducted study has some weaknesses and limitations which can be categorized as:

- Data. The effort of predicting academic success through data-centric investigations is frequently confronted with obstacles associated with the precision and dependability of the data. Critics might often highlight a multitude of concerns, involving possible situations of data manipulation and inherent biases present in experimental designs. Moreover, the constraints associated with datasets have a substantial influence on establishing the credibility of the study. It is worth noting that certain research studies may encounter limitations such as limited sample size or narrow scope, thereby inhibiting the broader applicability of their conclusions [10], [11], [33], [36], [37], [39], [52], [53], [57]. The inspection of data gathering methodology is frequently observed, particularly in cases where crucial variables are lacking. The absence of data presents a significant obstacle, which has the ability to introduce bias and distort the outcomes or understanding of a study [14], [61], [74]. There is a noticeable concern regarding the reliance of the majority of these studies reliance on a singular dataset, which raises inquiries regarding the wider applicability of their findings [17], [23], [80]. The exclusive emphasis on a specific aspect may unintentionally overlook essential factors that could have contributed to a more comprehensive analysis. Furthermore, it is important to acknowledge the presence of self-reporting and self-selection biases, which have been explicitly recognized in certain studies, thereby amplifying concerns regarding the reliability of the data [79], [86]. The complexities highlight the necessity for thorough examination of data in scholarly studies on academic performance prediction.
- Model Limitations. The process of modeling within the context of academic research is a complex activity, as numerous studies recognize and address the difficulties and constraints it faces. A frequent trend becomes apparent regarding the limitations of certain models, including their incapacity to assess specific types of information, disregarding essential attributes, or the potential for excessive simplification [10], [24], [33], [35], [37], [38], [40], [52], [54], [57]. This issue is further compounded by the decisions made by the researchers, such as the selection of the number of clusters in the algorithm, which can intrinsically influence the results of the study. Narrowing down the scope of research to a specific area, such as prioritizing a distant e-learning platform or employing a particular data gathering approach, may restrict the broader applicability of the research outcomes [77], [78], [84]. Several investigations have shown deficiencies in the modeling methodologies employed, encompassing algorithmic decisions and a lack of model interpretability. This raises issues over the potential for over-reliance on specific features and an excessive emphasis on certain measurements, which may lead to a biased interpretation of results [13], [15], [60], [72]. The unreadable character of numerous contemporary modeling methodologies presents difficulties, hindering comprehension and confidence in the results [12], [34], [56], [61]. Another problem lies in the verification of the applicability of these models in real-world situations, extending beyond controlled laboratory settings or specific academic courses. One significant issue that arises is the reliance of numerous studies on certain software or tools, which might introduce potential limitations [77]. The aforementioned issues collectively underscore the significance of consistently validating, optimizing, and scrutinizing the models employed in research. The diversity of educational contexts is a significant challenge in devising a universally applicable paradigm. Models trained on data from a single institution may not exhibit strong generalization capabilities to other institutions because of variations in teaching methodologies, grading systems, and other contextual data. Models established in one particular context may not effectively apply to a distict setting. Transferring a model that has been trained using data from one university to another may result in less ideal predictions due to variations in context.
- External Factors and Considerations. The discipline of academic performance prediction is substantially influenced by several external elements and considerations. While a considerable body of research focuses on the predicted accuracy of various models, there is a noticeable deficiency in explanatory analysis since many of these models neglect to thoroughly investigate the underlying causes that influence the outcomes [36], [60], [76]. The incorporation of novel methodologies into established systems frequently presents difficulties, underscoring the importance of harmonious cooperation among a wide range of participants [37], [58]. The significance of ethical considerations is of greatest significance, particularly in the implementation of AI and machine learning methodologies. It has been found that several studies have failed to adequately address concerns related to privacy and consent [1], [72], [73], [79]. The comprehensive examination of various influential elements of student performance continues to present a significant challenge for numerous studies. Furthermore, the issue of data gathering is further aggravated by ongoing obstacles such as ethical considerations, budgetary limitations, and several other restrictions [1], [79]. There is growing concern regarding the possible unintended consequences of predictions or interventions on students, as seen by the anticipation surrounding the potential correlation of course failures with the risk of dropping out [22], [81], [82], [85]. The process of applying research findings to practical education environments presents distinct difficulties, particularly when evaluating the potential

consequences associated with the adoption of specific tools or models [22], [81]. The inherent complexity and lack of transparency in many modeling methodologies exacerbate these challenges, restricting the ability to evaluate results accurately and fostering a sense of distrust. It is evident that, in addition to data and modeling, a wider array of external factors significantly influence the outcomes of academic research [60], [76].

By taking into consideration the aforementioned limitations, it is evident that there exist certain works that hold potential for further development by future researchers, which can be classified into the following categories:

- Data Preprocessing and Feature Engineering. To enhance the precision of prediction, attention is directed towards many factors, including academic background and emotional conditions. This can facilitate the development of a comprehensive understanding of the students' performance while also enhancing the accessibility and user-friendliness of the LMS [9], [24], [33], [80], [81].
- Model Adaptability. Encourages the development of flexible models that can be easily applied to a variety of educational environments in the real world. The focus of this review is on the deployment of practical and ethical considerations such as algorithmic fairness and a descriptive analysis of factors affecting student performance. The importance of developing ways to maintain and make the necessary transition from laboratory settings to those found in real-world environments is also emphasized [9], [24], [33], [36], [38], [60].
- Generalizability. This model undergoes testing on a range of datasets that increase in size and complexity, allowing for an assessment of its overall applicability. This study underscores the importance of verifying prediction models across multiple situations [9], [24], [38], [64].
- Explainability and Interpretability. This facilitates the comprehension and interpretation of models by educational stakeholders across diverse contexts. The utter importance of ensuring interpretability and doing rigorous testing of machine learning models is widely acknowledged. Additionally, it entails a willingness to disseminate models and data, thereby facilitating the active participation of community members [65], [67].
- Legal and Ethical Considerations. The significance of highlighting the necessity to take into account legal and ethical considerations arises when addressing issues concerning compliance with the General Data Protection Regulation (GDPR), acquiring informed permission, safeguarding data privacy, and preserving data security [18], [21], [67], [84]. The exploitation of student data. Which frequently encompasses confidential information, gives rise to concerns over privacy. Ensuring the confidentiality of student records is essential for adhering to privacy standards and upholding trust.
- Interventions. The proposal entails the development of interventions that can maximize their effectiveness through the utilization of predictive data [13], [62], [83], [86].
- Multidisciplinary Approach. To optimize the precision and efficacy of prediction models, it is strongly advised to incorporate valuable insights from a range of disciplines, including psychology. By integrating psychological principles into the process of constructing and enhancing these models, it is possible to attain predictions that are more resilient and dependable. The utilization of a cross-disciplinary approach enables us to access the information and comprehension acquired from the field of psychology in order to enhance our understanding of human behavior and the processes involved in decision-making. By utilizing these insights, we are able to gain a more profound comprehension of variables that impact outcomes, allowing us to enhance our predictive models and increase their comprehensiveness and accuracy. By adopting the predictive capacities of our models, leading to enhanced decision-making across diverse domains [63], [65].

IV. LIMITATIONS AND FUTURE WORKS

A. LIMITATIONS

Regardless of the extensive attempts that have been made, it is acknowledged that this review comes with certain limitations. Initially, it is worth noting that the chosen publications exclusively encompass works written in English, excluding a considerable body of research conducted in other languages. The inclusion of publications that utilize languages other than the primary language is deemed ineffective; however, this approach may limit the potential for selecting publications that make substantial contributions in this area of study. Furthermore, it is important to note that the research, particularly the stage involving the selection of studies, was conducted in March 2023. Consequently, any research regarding the prediction of academic performance that was published subsequent to March 2023 was not incorporated into the chosen publications.

B. FUTURE WORKS

The utilization of machine learning in forecasting student graduation has a significant potential for enhancing educational results. Nevertheless, there are certain domains that necessitate additional exploration in order to fully unleash the capabilities of this technology.

Methodological enhancements are required to augment the precision and dependability of predictive models. This entails sophisticated data processing and feature engineering methodologies, integrating students' emotional states and academic backgrounds, and constructing adaptive models that can proficiently operate in various educational settings. Furthermore, it is crucial to guarantee the generalization of the model by conducting tests on extensive and varied datasets in order to facilitate precise predictions in different scenarios. Moreover, prioritizing the interpretability of the

model will enhance the comprehension and confidence of stakeholders in the projections.

Investigating novel data sources is an additional crucial domain for future research. This involves examining unconventional data sources such as social media activity, online learning behavior, and wearable devices that might offer more profound insights into student engagement and learning habits. By incorporating new data sources with conventional academic data, researchers can construct more detailed and precise predictive models.

The emergence of machine learning techniques has promising prospects for advancing the area. Researchers might investigate the implementation of novel algorithms, advanced deep learning architectures, and ensemble methods to enhance the accuracy of predictions. Moreover, prioritizing the utilization of explainable AI techniques will provide educators with the ability to comprehend the underlying reasoning behind model predictions. This will empower them to make well-informed decisions regarding interventions and support strategies.

It is imperative to address ethical considerations at every stage of the research process. This encompasses guaranteeing the confidentiality and protection of data, acquiring explicit consent, and complying with applicable rules such as GDPR. Additionally, it is crucial to advocate for algorithmic fairness in order to prevent prejudiced forecasts that put certain students at a disadvantage.

Employing multidisciplinary methodologies is essential in order to create predictive models that are both efficient and influential. The cooperation of researchers in the fields of education, computer science, psychology, and other related disciplines will promote a thorough comprehension of student learning and expedite the creation of strong and widely applicable models.

Future research in the field of machine learning for predicting student graduation should aim beyond the mere task of outcome prediction. The goal should be to revolutionize educational methods by equipping educators with practical knowledge to customize learning experiences, detect students at risk of failure at an early stage, and apply specific interventions that foster student achievement. Researchers may develop a dynamic and transformative approach that promotes student achievement and enhances educational results for all by efficiently combining advanced methods, analyzing various data sources, and utilizing emerging machine learning approaches.

V. CONCLUSION

The findings of the review yielded many conclusions. Related to RQ1, a significant portion of the scholarly investigations conducted in the field of educational data mining and educational data analytics, particularly those focused on higher education, pertain to the prediction of academic performance, the identification of students at risk, the anticipation of dropout rates or graduation rates, and the analysis of student engagement.

In response to RQ2, the primary data resources utilized in this study are predominantly LMS and SIS. The data frequently employed for predictive purposes encompasses academic, behavioral, demographic, pre-university, and university entrance examination data. This research revealed that LMS and SIS are the predominant data sources utilized. Upon further investigation, it was discovered that this was due to the data generated by the two data sources. LMS and SIS provide extensive data pertaining to students' academic progression. LMS houses information regarding students' engagement with digital educational resources, tasks, and evaluations, whereas SIS primarily retains demographic data, enrollment particulars, and academic records. Both systems provide longitudinal data, enabling machine learning algorithms to analyze the progression of students over a period of time. The longitudinal approach is especially valuable for predicting academic outcomes and detecting trends or patterns that could impact student achievement. Both data sources are essential elements of educational institutions, facilitating convenient access to the data they contain. This accessibility enables the deployment of machine learning models without substantial obstacles in data procurement. The SVM algorithm has been extensively employed in a total of 16 publications, making it the technique with the highest frequency of utilization. In a total of 14 studies, the RF technique was employed subsequent to the utilization of SVM. Additionally, LR was employed in 9 papers. In addition, the study employed the k-NN algorithm and the ANN/MLP model. The study revealed that SVM, RF, and LR have become common algorithms for predicting student academic performance. SVM features a parameter for regularization that aids managing overfitting, which is essential in dealing with a small number of data in student performance prediction tasks [37]. All three methods have a track record of success in addressing past educational prediction challenges and are capable of handling both binary and multi-class classification issues. Furthermore, these algorithms have customizable parameters that may be fine-tuned to optimize performance for specific datasets. The flexibility of these models enables academics and data scientists to optimize them based on the specific attributes of the educational data being utilized. The methods can be adjusted to accommodate datasets of any magnitude. Given the wide range of student numbers and characteristics in educational datasets, algorithms capable of accommodating diverse scales are preferred.

RQ3 pertains to the identification of research gaps and challenges within the study. These include limited data availability, incomplete data resulting from inadequate data collection methods, dependence on a single data source, limitations of certain models that hinder their ability to assess specific types of information, disregard for important studies, potential oversimplification, and an excessive reliance on particular features or measurements, which may introduce bias into the interpretation of results. The significance of ethical issues has utmost relevance, particularly in

the utilization of AI and machine learning approaches. There is an increasing apprehension over the potential inadvertent ramifications of forecasting or intervention on students.

Smart learning environment, particularly learners module, can benefit from this study by acknowledging the widespread use of LMS and SIS as key data sources. Specifically, the learners module can leverage academic, behavioral, demographic, pre-university and university admission data to make predictions. Additionally, information on SVM as the most commonly used method, followed by RF, LR, k-NN, and ANN/MLP, can assist learners module in selecting or improving machine learning methods that can be employed. Moreover, educational policies can conduct more informed and comprehensive interventions based on the results obtained from learners module.

REFERENCES

[1] S. N. Liao, D. Zingaro, K. Thai, C. Alvarado, W. G. Griswold, and L. Porter, "A robust machine learning technique to predict low-performing students," *ACM Trans. Comput. Educ.*, vol. 19, no. 3, pp. 1–19, Sep. 2019, doi: 10.1145/3277569.

[2] N. M. Suhaimi, S. Abdul-Rahman, S. Mutalib, N. H. A. Hamid, and A. Hamid, "Review on predicting students' graduation time using machine learning algorithms," *Int. J. Modern Educ. Comput. Sci.*, vol. 11, no. 7, pp. 1–13, Jul. 2019, doi: 10.5815/ijmecs.2019.07.01.

[3] C. G. Serrano and J. A. Mosquera-Bolaños, "Leadership 5.0. a new approach in higher education," *IEEE Rev. Iberoam. Tecnol. Aprendiz.*, vol. 17, no. 4, pp. 393–400, Nov. 2022, doi: 10.1109/RITA.2022.3217195.

[4] S. Ranjeeth, T. P. Latchoumi, and P. V. Paul, "A survey on predictive models of learning analytics," *Proc. Comput. Sci.*, vol. 167, pp. 37–46, Mar. 2020, doi: 10.1016/j.procs.2020.03.180.

[5] E. Alyahyan and D. Düştegör, "Predicting academic success in higher education: Literature review and best practices," *Int. J. Educ. Technol. Higher Educ.*, vol. 17, no. 1, p. 3, Dec. 2020, doi: 10.1186/s41239-020-0177-7.

[6] Y.-H. Hu, C.-L. Lo, and S.-P. Shih, "Developing early warning systems to predict students' online learning performance," *Comput. Hum. Behav.*, vol. 36, pp. 469–478, Jul. 2014, doi: 10.1016/j.chb.2014.04.002.

[7] S. U. Khan, S. A. K. Bangash, and K. U. Khan, "Learning analytics in the era of big data: A systematic literature review protocol," in *Proc. Int. Symp. Wireless Syst. Netw. (ISWSN)*, Nov. 2017, pp. 1–7, doi: 10.1109/ISWSN.2017.8250033.

[8] S. G. Essa, T. Celik, and N. E. Human-Hendricks, "Personalized adaptive learning technologies based on machine learning techniques to identify learning styles: A systematic literature review," *IEEE Access*, vol. 11, pp. 48392–48409, 2023, doi: 10.1109/ACCESS.2023.3276439.

[9] J. Figueroa-Cañas and T. Sancho-Vinuesa, "Early prediction of dropout and final exam performance in an online statistics course," *IEEE Rev. Iberoam. Tecnol. Aprendiz.*, vol. 15, no. 2, pp. 86–94, May 2020, doi: 10.1109/RITA.2020.2987727.

[10] A. Gupta, D. Garg, and P. Kumar, "Mining sequential learning trajectories with hidden Markov models for early prediction of at-risk students in e-learning environments," *IEEE Trans. Learn. Technol.*, vol. 15, no. 6, pp. 783–797, Dec. 2022, doi: 10.1109/TLT.2022.3197486.

[11] A. Akram, C. Fu, Y. Li, M. Y. Javed, R. Lin, Y. Jiang, and Y. Tang, "Predicting students' academic procrastination in blended learning course using homework submission data," *IEEE Access*, vol. 7, pp. 102487–102498, 2019, doi: 10.1109/ACCESS.2019.2930867.

[12] S. Qu, K. Li, S. Zhang, and Y. Wang, "Predicting achievement of students in smart campus," *IEEE Access*, vol. 6, pp. 60264–60273, 2018, doi: 10.1109/ACCESS.2018.2875742.

[13] R. Alcaraz, A. Martínez-Rodrigo, R. Zangróniz, and J. J. Rieta, "Early prediction of students at risk of failing a face-to-face course in power electronic systems," *IEEE Trans. Learn. Technol.*, vol. 14, no. 5, pp. 590–603, Oct. 2021, doi: 10.1109/TLT.2021.3118279.

[14] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, and P. C. Nshimyumukiza, "Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization," *Comput. Educ. Artif. Intell.*, vol. 3, no. 100066, 2022, doi: 10.1016/j.caeai.2022.100066.

[15] J. A. Talamás-Carvajal and H. G. Ceballos, "A stacking ensemble machine learning method for early identification of students at risk of dropout," *Educ. Inf. Technol.*, vol. 28, no. 9, pp. 12169–12189, Mar. 2023, doi: 10.1007/s10639-023-11682-z.

[16] Y. Ma, C. Cui, J. Yu, J. Guo, G. Yang, and Y. Yin, "Multi-task MIML learning for pre-course student performance prediction," *Frontiers Comput. Sci.*, vol. 14, no. 5, Oct. 2020, Art. no. 145313, doi: 10.1007/s11704-019-9062-8.

[17] S. Alturki, L. Cohausz, and H. Stuckenschmidt, "Predicting master's students' academic performance: An empirical study in Germany," *Smart Learn. Environments*, vol. 9, no. 1, p. 38, Dec. 2022, doi: 10.1186/s40561-022-00220-y.

[18] H. Yao, D. Lian, Y. Cao, Y. Wu, and T. Zhou, "Predicting academic performance for college students: A campus behavior perspective," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 3, pp. 1–21, May 2019, doi: 10.1145/3299087.

[19] A. Jokhan, B. Sharma, and S. Singh, "Early warning system as a predictor for student performance in higher education blended courses," *Stud. Higher Educ.*, vol. 44, no. 11, pp. 1900–1911, Nov. 2019, doi: 10.1080/03075079.2018.1466872.

[20] M. Cannistrà, C. Masci, F. Ieva, T. Agasisti, and A. M. Paganoni, "Early-predicting dropout of university students: An application of innovative multilevel machine learning and statistical techniques," *Stud. Higher Educ.*, vol. 47, no. 9, pp. 1935–1956, Sep. 2022, doi: 10.1080/03075079.2021.2018415.

[21] O. Moscoso-Zea, P. Saa, and S. Luján-Mora, "Evaluation of algorithms to predict graduation rate in higher education institutions by applying educational data mining," *Australas. J. Eng. Educ.*, vol. 24, no. 1, pp. 4–13, Jan. 2019, doi: 10.1080/22054952.2019.1601063.

[22] J. M. Trussel and L. Burke-Smalley, "Demography and student success: Early warning tools to drive intervention," *J. Educ. Bus.*, vol. 93, no. 8, pp. 363–372, Nov. 2018, doi: 10.1080/08832323.2018.1496893.

[23] P. Everaert, E. Opdecam, and H. van der Heijden, "Predicting first-year university progression using early warning signals from accounting education: A machine learning approach," *Accounting Educ.*, vol. 33, no. 1, pp. 1–26, Nov. 2022, doi: 10.1080/09639284.2022.2145570.

[24] L. Zhao, K. Chen, J. Song, X. Zhu, J. Sun, B. Caulfield, and B. M. Namee, "Academic performance prediction based on multisource, multifeature behavioral data," *IEEE Access*, vol. 9, pp. 5453–5465, 2021, doi: 10.1109/ACCESS.2020.3002791.

[25] N. L. Wade, "Measuring, manipulating, and predicting student success: A 10-year assessment of carnegie R1 doctoral universities between 2004 and 2013," *J. College Student Retention, Res., Theory Pract.*, vol. 21, no. 1, pp. 119–141, May 2019, doi: 10.1177/1521025119831456.

[26] C. Baek and T. Doleck, "Educational data mining: A bibliometric analysis of an emerging field," *IEEE Access*, vol. 10, pp. 31289–31296, 2022, doi: 10.1109/ACCESS.2022.3160457.

[27] *Educationaldatamining*. Accessed: Jul. 13, 2023. [Online]. Available: https://educationaldatamining.org/

[28] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Educ. Inf. Technol.*, vol. 28, no. 1, pp. 905–971, Jan. 2023, doi: 10.1007/s10639-022-11152-y.

[29] B. Prenkaj, P. Velardi, G. Stilo, D. Distante, and S. Faralli, "A survey of machine learning approaches for student dropout prediction in online courses," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–34, May 2021, doi: 10.1145/3388792.

[30] C. Korkmaz and A.-P. Correia, "A review of research on machine learning in educational technology," *Educ. Media Int.*, vol. 56, no. 3, pp. 250–267, Jul. 2019, doi: 10.1080/09523987.2019.1669875.

[31] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student' performance prediction using machine learning techniques," *Educ. Sci.*, vol. 11, no. 9, p. 552, Sep. 2021, doi: 10.3390/educsci11090552.

[32] B. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering," *Inf. Softw. Technol.*, vol. 55, no. 12, pp. 2049–2075, Dec. 2013, doi: 10.1016/j.infsof.2013.07.010.

[33] G. Czibula, G. Ciubotariu, M.-I. Maier, and H. Lisei, "IntelliDaM: A machine learning-based framework for enhancing the performance of decision-making processes. A case study for educational data mining," *IEEE Access*, vol. 10, pp. 80651–80666, 2022, doi: 10.1109/ACCESS.2022.3195531.

[34] M. M. Rahman, Y. Watanobe, T. Matsumoto, R. U. Kiran, and K. Nakamura, "Educational data mining to support programming learning using problem-solving data," *IEEE Access*, vol. 10, pp. 26186–26202, 2022, doi: 10.1109/ACCESS.2022.3157288.

[35] M. Á. Prada, M. Domínguez, J. L. Vicario, P. A. V. Alves, M. Barbu, M. Podpora, U. Spagnolini, M. J. V. Pereira, and R. Vilanova, "Educational data mining for tutoring support in higher education: A web-based tool case study in engineering degrees," *IEEE Access*, vol. 8, pp. 212818–212836, 2020, doi: 10.1109/ACCESS.2020.3040858.

[36] A. Alshanqiti and A. Namoun, "Predicting student performance and its influential factors using hybrid regression and multi-label classification," *IEEE Access*, vol. 8, pp. 203827–203844, 2020, doi: 10.1109/ACCESS.2020.3036572.

[37] L. N. Singelmann and D. L. Ewert, "Leveraging the innovation-based learning framework to predict and understand student success in innovation," *IEEE Access*, vol. 10, pp. 36123–36139, 2022, doi: 10.1109/ACCESS.2022.3163744.

[38] D. Hooshyar, Y. Yang, M. Pedaste, and Y.-M. Huang, "Clustering algorithms in an educational context: An automatic comparative approach," *IEEE Access*, vol. 8, pp. 146994–147014, 2020, doi: 10.1109/ACCESS.2020.3014948.

[39] S. López-Pernas and M. Saqr, "Bringing synchrony and clarity to complex multi-channel data: A learning analytics study in programming education," *IEEE Access*, vol. 9, pp. 166531–166541, 2021, doi: 10.1109/ACCESS.2021.3134844.

[40] A. Rafique, M. S. Khan, M. H. Jamal, M. Tasadduq, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, "Integrating learning analytics and collaborative learning for improving student's academic performance," *IEEE Access*, vol. 9, pp. 167812–167826, 2021, doi: 10.1109/ACCESS.2021.3135309.

[41] H. E. Abdelkader, A. G. Gad, A. A. Abohany, and S. E. Sorour, "An efficient data mining technique for assessing satisfaction level with online learning for higher education students during the COVID-19," *IEEE Access*, vol. 10, pp. 6286–6303, 2022, doi: 10.1109/ACCESS.2022.3143035.

[42] P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Muñoz-Merino, I. Estévez-Ayres, and C. D. Kloos, "A learning analytics methodology for understanding social interactions in MOOCs," *IEEE Trans. Learn. Technol.*, vol. 12, no. 4, pp. 442–455, Oct. 2019, doi: 10.1109/TLT.2018.2883419.

[43] P. M. Moreno-Marcos, T.-C. Pong, P. J. Muñoz-Merino, and C. D. Kloos, "Analysis of the factors influencing learners' performance prediction with learning analytics," *IEEE Access*, vol. 8, pp. 5264–5282, 2020, doi: 10.1109/ACCESS.2019.2963503.

[44] F. A. S. Herrera, R. G. Crespo, L. R. Baena, and D. Burgos, "A solution to manage the full life cycle of learning analytics in a learning management system: AnalyTIC," *IEEE Rev. Iberoam. Tecnol. Aprendiz.*, vol. 14, no. 4, pp. 127–134, Nov. 2019, doi: 10.1109/RITA.2019.2950148.

[45] S. Kausar, X. Huahu, I. Hussain, Z. Wenhao, and M. Zahid, "Integration of data mining clustering approach in the personalized e-learning system," *IEEE Access*, vol. 6, pp. 72724–72734, 2018, doi: 10.1109/ACCESS.2018.2882240.

[46] G. Kostopoulos, S. Karlos, and S. Kotsiantis, "Multiview learning for early prognosis of academic performance: A case study," *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 212–224, Apr. 2019, doi: 10.1109/TLT.2019.2911581.

[47] L. Sha, M. Rakovic, A. Das, D. Gaşevic, and G. Chen, "Leveraging class balancing techniques to alleviate algorithmic bias for predictive tasks in education," *IEEE Trans. Learn. Technol.*, vol. 15, no. 4, pp. 481–492, Aug. 2022, doi: 10.1109/TLT.2022.3196278.

[48] L. Gao, Z. Zhao, L. Qi, Y. Liang, and J. Du, "Modeling the effort and learning ability of students in MOOCs," *IEEE Access*, vol. 7, pp. 128035–128042, 2019, doi: 10.1109/ACCESS.2019.2937985.

[49] S. Alwarthan, N. Aslam, and I. U. Khan, "An explainable model for identifying at-risk student at higher education," *IEEE Access*, vol. 10, pp. 107649–107668, 2022, doi: 10.1109/ACCESS.2022.3211070.

[50] Y. Qu, F. Li, L. Li, X. Dou, and H. Wang, "Can we predict student performance based on tabular and textual data?" *IEEE Access*, vol. 10, pp. 86008–86019, 2022, doi: 10.1109/ACCESS.2022.3198682.

[51] K. Liu, S. Tatinati, and A. W. H. Khong, "Context-based data model for effective real-time learning analytics," *IEEE Trans. Learn. Technol.*, vol. 13, no. 4, pp. 790–803, Oct. 2020, doi: 10.1109/TLT.2020.3027441.

[52] M. Liz-Domínguez, M. Llamas-Nistal, M. Caeiro-Rodríguez, and F. A. Mikic-Fonte, "Profiling students' self-regulation with learning analytics: A proof of concept," *IEEE Access*, vol. 10, pp. 71899–71913, 2022, doi: 10.1109/ACCESS.2022.3187732.

[53] R. Boegeholz, J. Guerra, and E. Scheihing, "Exploring risk of delay in academic trajectories in two undergraduate programs," *IEEE Rev. Iberoam. Tecnol. Aprendiz.*, vol. 17, no. 3, pp. 290–300, Aug. 2022, doi: 10.1109/RITA.2022.3191298.

[54] H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," *IEEE Access*, vol. 8, pp. 55462–55470, 2020, doi: 10.1109/ACCESS.2020.2981905.

[55] R. Ghorbani and R. Ghousi, "Comparing different resampling methods in predicting students' performance using machine learning techniques," *IEEE Access*, vol. 8, pp. 67899–67911, 2020, doi: 10.1109/ACCESS.2020.2986809.

[56] W. Xing, D. Du, A. Bakhshi, K.-C. Chiu, and H. Du, "Designing a transferable predictive model for online learning using a Bayesian updating approach," *IEEE Trans. Learn. Technol.*, vol. 14, no. 4, pp. 474–485, Aug. 2021, doi: 10.1109/TLT.2021.3107349.

[57] P. Krieter, "Are you still there? An exploratory case study on estimating students' LMS online time by combining log files and screen recordings," *IEEE Trans. Learn. Technol.*, vol. 15, no. 1, pp. 55–63, Feb. 2022, doi: 10.1109/TLT.2022.3154828.

[58] A. Cano and J. D. Leonard, "Interpretable multiview early warning system adapted to underrepresented student populations," *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 198–211, Apr. 2019, doi: 10.1109/TLT.2019.2911079.

[59] Y. Wen, Y. Tian, B. Wen, Q. Zhou, G. Cai, and S. Liu, "Consideration of the local correlation of learning behaviors to predict dropouts from MOOCs," *Tsinghua Sci. Technol.*, vol. 25, no. 3, pp. 336–347, Jun. 2020, doi: 10.26599/TST.2019.9010013.

[60] A. Ortigosa, R. M. Carro, J. Bravo-Agapito, D. Lizcano, J. J. Alcolea, and Ó. Blanco, "From lab to production: Lessons learnt and real-life challenges of an early student-dropout prevention system," *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 264–277, Apr./Jun. 2019, doi: 10.1109/TLT.2019.2911608.

[61] H. Prabowo, A. A. Hidayat, T. W. Cenggoro, R. Rahutomo, K. Purwandari, and B. Pardamean, "Aggregating time series and tabular data in deep learning model for university students' GPA prediction," *IEEE Access*, vol. 9, pp. 87370–87377, 2021, doi: 10.1109/ACCESS.2021.3088152.

[62] G. Deeva, J. De Smedt, and J. De Weerdt, "Educational sequence mining for dropout prediction in MOOCs: Model building, evaluation, and benchmarking," *IEEE Trans. Learn. Technol.*, vol. 15, no. 6, pp. 720–735, Dec. 2022, doi: 10.1109/TLT.2022.3215598.

[63] E. Popescu and F. Leon, "Predicting academic performance based on learner traces in a social learning environment," *IEEE Access*, vol. 6, pp. 72774–72785, 2018, doi: 10.1109/ACCESS.2018.2882297.

[64] D. M. Olivé, D. Q. Huynh, M. Reynolds, M. Dougiamas, and D. Wiese, "A quest for a one-size-fits-all neural network: Early prediction of students at risk in online courses," *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 171–183, Apr. 2019, doi: 10.1109/TLT.2019.2911068.

[65] J.-L. Hung, B. E. Shelton, J. Yang, and X. Du, "Improving predictive modeling for at-risk student identification: A multistage approach," *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 148–157, Apr. 2019, doi: 10.1109/TLT.2019.2911072.

[66] K. Mangaroska, B. Vesin, V. Kostakos, P. Brusilovsky, and M. N. Giannakos, "Architecting analytics across multiple e-learning systems to enhance learning design," *IEEE Trans. Learn. Technol.*, vol. 14, no. 2, pp. 173–188, Apr. 2021, doi: 10.1109/TLT.2021.3072159.

[67] M. Saarela, V. Heilala, P. Jääskelä, A. Rantakaulio, and T. Kärkkäinen, "Explainable student agency analytics," *IEEE Access*, vol. 9, pp. 137444–137459, 2021, doi: 10.1109/ACCESS.2021.3116664.

[68] J. E. Yoo, M. Rho, and Y. Lee, "Online students' learning behaviors and academic success: An analysis of LMS log data from flipped classrooms via regularization," *IEEE Access*, vol. 10, pp. 10740–10753, 2022, doi: 10.1109/ACCESS.2022.3144625.

[69] M. M. Rahman, Y. Watanobe, R. U. Kiran, T. C. Thang, and I. Paik, "Impact of practical skills on academic performance: A data-driven analysis," *IEEE Access*, vol. 9, pp. 139975–139993, 2021, doi: 10.1109/ACCESS.2021.3119145.

[70] M. Riestra-González, M. D. P. Paule-Ruíz, and F. Ortin, "Massive LMS log data analysis for the early prediction of course-agnostic student performance," *Comput. Educ.*, vol. 163, Apr. 2021, Art. no. 104108, doi: 10.1016/j.compedu.2020.104108.
[71] I. E. Livieris, T. Kotsilieris, V. Tampakas, and P. Pintelas, "Improving the evaluation process of students' performance utilizing a decision support software," *Neural Comput. Appl.*, vol. 31, no. 6, pp. 1683–1694, Jun. 2019, doi: 10.1007/s00521-018-3756-y.
[72] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using machine learning to predict student difficulties from learning session data," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 381–407, Jun. 2019, doi: 10.1007/s10462-018-9620-8.
[73] P. Jiao, F. Ouyang, Q. Zhang, and A. H. Alavi, "Artificial intelligence-enabled prediction model of student academic performance in online engineering education," *Artif. Intell. Rev.*, vol. 55, no. 8, pp. 6321–6344, Dec. 2022, doi: 10.1007/s10462-022-10155-y.
[74] L. Ramanathan, G. Parthasarathy, K. Vijayakumar, L. Lakshmanan, and S. Ramani, "Cluster-based distributed architecture for prediction of student's performance in higher education," *Cluster Comput.*, vol. 22, no. 1, pp. 1329–1344, Jan. 2019, doi: 10.1007/s10586-017-1624-7.
[75] J. Kuzilek, Z. Zdrahal, J. Vaclavek, V. Fuglik, J. Skocilas, and A. Wolff, "First-year engineering students' strategies for taking exams," *Int. J. Artif. Intell. Educ.*, vol. 33, no. 3, pp. 583–608, Aug. 2022, doi: 10.1007/s40593-022-00303-4.
[76] G. Akçapınar, M. N. Hasnine, R. Majumdar, B. Flanagan, and H. Ogata, "Developing an early-warning system for spotting at-risk students by using eBook interaction logs," *Smart Learn. Environments*, vol. 6, no. 1, p. 4, Dec. 2019, doi: 10.1186/s40561-019-0083-4.
[77] M. Yağcı, "Educational data mining: Prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 1, p. 11, Dec. 2022, doi: 10.1186/s40561-022-00192-z.
[78] G. Akçapınar, A. Altun, and P. Aşkar, "Using learning analytics to develop early-warning system for at-risk students," *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, p. 40, Dec. 2019, doi: 10.1186/s41239-019-0172-z.
[79] A.-E. Guerrero-Roldán, M. E. Rodríguez-González, D. Bañeres, A. Elasri-Ejjaberi, and P. Cortadas, "Experiences in the use of an adaptive intelligent system to enhance online learners' performance: A case study in economics and business courses," *Int. J. Educ. Technol. Higher Educ.*, vol. 18, no. 1, p. 36, Dec. 2021, doi: 10.1186/s41239-021-00271-0.
[80] R. Bertolini, S. J. Finch, and R. H. Nehm, "Enhancing data pipelines for forecasting student performance: Integrating feature selection with cross-validation," *Int. J. Educ. Technol. Higher Educ.*, vol. 18, no. 1, p. 44, Dec. 2021, doi: 10.1186/s41239-021-00279-6.
[81] B. Albreiki, T. Habuza, and N. Zaki, "Framework for automatically suggesting remedial actions to help students at risk based on explainable ML and rule-based models," *Int. J. Educ. Technol. Higher Educ.*, vol. 19, no. 1, p. 49, Sep. 2022, doi: 10.1186/s41239-022-00354-6.
[82] D. Bañeres, M. E. Rodríguez-González, A.-E. Guerrero-Roldán, and P. Cortadas, "An early warning system to identify and intervene online dropout learners," *Int. J. Educ. Technol. Higher Educ.*, vol. 20, no. 1, p. 3, Jan. 2023, doi: 10.1186/s41239-022-00371-5.
[83] F. Ouyang, M. Wu, L. Zheng, L. Zhang, and P. Jiao, "Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course," *Int. J. Educ. Technol. Higher Educ.*, vol. 20, no. 1, p. 4, Jan. 2023, doi: 10.1186/s41239-022-00372-4.
[84] A. V. Bengesai and V. Paideya, "An analysis of academic and institutional factors affecting graduation among engineering students at a South African University," *Afr. J. Res. Math., Sci. Technol. Educ.*, vol. 22, no. 2, pp. 137–148, May 2018, doi: 10.1080/18117295.2018.1456770.
[85] M. Chen, X. Huang, H. Chen, X. Su, and J. Yur-Austin, "Data driven course scheduling to ensure timely graduation," *Int. J. Prod. Res.*, vol. 61, no. 1, pp. 336–361, Jan. 2023, doi: 10.1080/00207543.2021.1916118.
[86] T. Goad, E. Jones, S. Bulger, D. Daum, N. Hollett, and E. Elliott, "Predicting student success in online physical education," *Amer. J. Distance Educ.*, vol. 35, no. 1, pp. 17–32, Jan. 2021, doi: 10.1080/08923647.2020.1829254.
[87] N. Sghir, A. Adadi, and M. Lahmer, "Recent advances in predictive learning analytics: A decade systematic review (2012–2022)," *Educ. Inf. Technol.*, vol. 28, no. 7, pp. 8299–8333, Jul. 2023, doi: 10.1007/s10639-022-11536-0.

LIDYA R. PELIMA received the bachelor's degree in informatics engineering from Atma Jaya University, Yogyakarta, in 2007. She is currently pursuing the master's degree in electrical engineering with the Bandung Institute of Technology (ITB). Her current research interests include educational technology and machine learning.

YUDA SUKMANA received the bachelor's degree in electrical engineering education from the Indonesia University of Education (UPI), Bandung, Indonesia, and the master's degree in electrical engineering from the Bandung Institute of Technology (ITB), where he is currently pursuing the Ph.D. degree in electrical engineering and informatics. His research interests include educational technology, mobile applications, and artificial intelligence.

YUSEP ROSMANSYAH received the bachelor's degree in electrical engineering from the Bandung Institute of Technology (ITB), Indonesia, in 1993, and the M.Sc. and Ph.D. degrees from the University of Surrey, U.K., in 1996 and 2003, respectively. He has been a Researcher and a Professor of smart multimedia processing with the School of Electrical Engineering and Informatics, ITB. His research interests include multimedia, educational technology, and cyber security.