

Students' Academic Performance Prediction Using Educational Data Mining and Machine Learning: A Systematic Review

Munaf Salim Najim Al-Din and Hussein Ali Al Abdulqader

Department of Electrical and Computer Engineering, College of Engineering and Architecture,
University of Nizwa, Nizwa, Oman

DOI: <https://dx.doi.org/10.47772/IJRISS.2024.808095>

Received: 18 August 2024; Accepted: 23 August 2024; Published: 31 August 2024

ABSTRACT

Forecasting the performance of students holds paramount importance in the context of higher education since the criteria for a high quality university is based on its excellent record of academic achievements. At the present time, predicting students' performance becomes more challenging due to the huge increase in the amount of educational data that is now available in educational databases. With the introduction of information systems and data mining and machine learning techniques in education a new era has been started to reveal the methodologies in studying and analyzing students' academic performance and to enable the recording and retention of large volumes of data in educational institutions. This paper seeks to systematically review the current research on predicting student performance through the use of educational data mining and machine learning techniques. The review synthesizes a wide range of studies, encompassing diverse educational levels, data sources, and predictive models. A comprehensive review was conducted for available research spanning from 2015 to 2023, to provide a foundational understanding of the intelligent methods employed in forecasting student performance. The search encompassed different electronic bibliographic databases, such as IEEE Xplore, Google Scholar, and Science Direct. In this paper, 17 survey papers and 74 research papers have been examined and analyzed, emphasizing seven key aspects that aim to have interpretable models for forecasting student performance.

Keywords: Students' Academic Performance, Educational Data Mining, Machine Learning

INTRODUCTION

Modern educational establishments function within an intensely competitive and intricate environments as a result of prevalent challenges, such as delivering education of exceptional quality, devising approaches to assess student achievement, and discerning forthcoming requirements. It is well known that educational institutes put in diligent effort to establish and enhance strategies that address students' challenges throughout their academic journey. One of the most important challenges is to continuously anticipating student achievements upon enrollment and as time progresses. These anticipation of students' performance serves as a valuable tool for aiding the selection of courses and devising tailored future study plans. In addition, they facilitate educators and administrators in monitoring students, enabling the provision of timely assistance and the alignment of training programs to optimize outcomes. As a consequence this aids these institutes in adeptly refining these strategies, benefiting both administrative staff and educators involved in the students' progress. Consequently, educational institutions strive to formulate individual student models that forecast each student's attributes and performance [1].

Over the past three decades, as computers have become increasingly ubiquitous, educational institutions have begun amassing extensive volumes of data. This rich reserve of data serves as the cornerstone of

educational data mining, a specialized field that leverages the capabilities of big data analytics to unearth valuable insights and patterns within the educational domain. Consequently, data mining has emerged as one of the most prevalent techniques employed to evaluate students' performance, finding extensive application within the educational sector, commonly referred to as Educational Data Mining (EDM) [2]. EDM is a method utilized to extract valuable information and discern patterns from vast educational databases, offering the potential to revolutionize decision-making processes for educators, improve curriculum design, and personalize teaching approaches to cater to the unique needs of individual students. The fusion of big data and educational data mining not only enhances educational outcomes but also assumes a pivotal role in shaping the future landscape of teaching and learning. EDM is, in essence, an interdisciplinary research field that amalgamates machine learning (ML), statistics, data mining (DM), educational psychology, cognitive psychology, and other theories and methodologies to dissect educational data and address a variety of challenges. The interdisciplinary nature of EDM is one of its strengths, as it allows researchers to draw on a variety of perspectives and methodologies to develop innovative solutions to educational problems. These disciplines include, Computer science, Education, Psychology, Statistics and Social science.

Typically, EDM focuses on data stored within educational Learning Management Systems (LMS) and databases. These methods are used to examine data generated during educational processes, revealing hidden insights, relationships, and patterns within vast data repositories. While multiple definitions of EDM exist, they share a common theme: it is a growing field dedicated to developing techniques specifically designed to analyze educational data, providing deeper insights into both students and their learning environments. As outlined in reference [3], the EDM cycle represents a customized adaptation of the general DM cycle, finely tuned to the intricacies of educational environments. Comprising four stages encompassing nine distinct steps, this cycle is designed to directly tackle the distinctive challenges and objectives prevalent in the field of education, as illustrated in Figure 1.

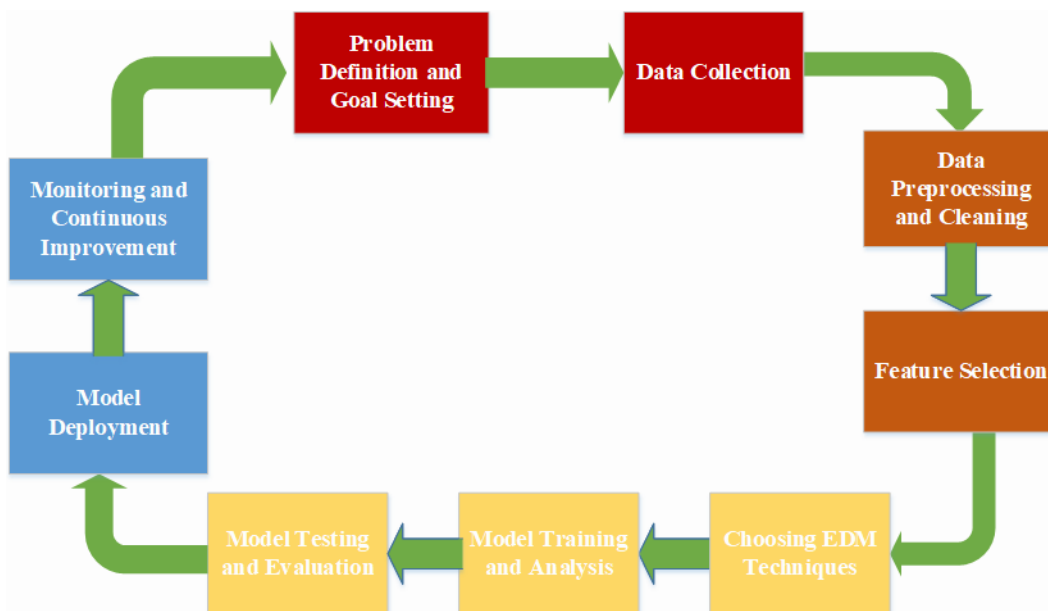


Fig. 1. Typical Education Data Mining Cycle.

The EDM process commences with clearly defining the problem and establishing goals. This entails precisely outlining the educational issue or question being investigated. During this phase, project objectives, goals, and the formulation of fundamental research questions occur. EDM researchers utilize various data mining techniques to perform a range of tasks that are broadly categorized into five groups namely; Regression, Classification, Clustering, Anomaly Detection and Recommendation. The second

stage, comprising Data Collection, Preparation, and Feature Selection, is the most time-intensive, taking up to 80% of the entire process. It involves gathering information from diverse educational sources, merging it into a unified dataset, addressing missing values and outliers, and selecting relevant features for analysis. The third stage involves three key steps: choosing EDM techniques, Model Training and Analysis, and Model Testing and Evaluation. In the first step, appropriate data mining and machine learning techniques are selected based on educational goals, mainly focusing on classification, clustering, prediction, and association of learning activities. In the second step, chosen techniques are applied to collected data using classical ML methods, generating insights for refining teaching methods and identifying at-risk students. The final step evaluates the models' effectiveness using educational metrics to enhance outcomes. The final stage consists of two main functions: model deployment and model monitoring and improvement. In the first function, insights and recommendations are implemented into educational practices, integrating data-driven decision-making into teaching strategies and institutional policies. In the second function, the performance of educational interventions is continuously monitored, and models are adapted and refined based on ongoing data and feedback from the educational environment.

To delve into the latest implementations of EDM techniques applied by researchers to forecast student performance, this investigation reviews some of the pertinent research in this field. The first objective of this paper is to present a structured evaluation of the present state of research in forecasting student performance through the utilization of EDM and ML techniques. This review consolidates an extensive array of investigations that encompass various educational levels, data origins, and predictive models. Secondly, to establish a solid foundation in understanding the intelligent techniques employed in predicting student achievement, a comprehensive examination of research conducted from 2015 to 2023 was carried out. This investigation encompassed a variety of electronic bibliographic databases, including IEEE Xplore, Google Scholar, and Science Direct.

LITERATURE REVIEW

Educational data mining involves the use of data analysis techniques to extract valuable insights from educational data. When applied to students' academic performance, researchers often aim to identify patterns, predictors, and interventions that can enhance learning outcomes. Consequently, the prediction of students' academic achievements has long captivated the attention of policymakers, researchers, and educators. As a result, numerous studies have been conducted to assess current approaches for predicting student performance using EDM. Considerable number of survey papers have been published on this topic; nevertheless, in this study 17 survey papers were selected as a starting point. These survey papers were carefully selected to analyze and identify appropriate methods for existing parameters, address gaps in current research, provide context for a new research initiative, and contribute to the development of a comprehensive and methodologically robust study. This contribution involves clarifying the research objectives, methodologies, and predictive variables utilized in the field. According to [4-5] the major aspects that need to be covered by a systematic review includes namely, Research Question, Objectives, Search Strategy, Inclusion and Exclusion Criteria, Study Selection Process, Data Extraction, Quality Assessment, Synthesis of Findings, Meta-Analysis (if applicable). Table I presents a comparison of the survey papers analyzed in this study which includes some of the aforementioned aspects.

Based on the survey papers listed in table I, the majority of contributors to existing surveys employed the systematic literature survey method to analyze various aspects of prior studies, including publication year, key contributors, the source of experimental datasets, and more. Numerous studies were categorized based on the EDM methods applied or the issues addressed in the education domain. The primary aim of the authors is to analyze and extract the objectives and research questions from each article, with the intention of formulating a comprehensive objective and research question for this study. This overarching goal aims to address various aspects concerning the application of EDM and ML in predicting students' academic

performance.

In addition to the main intention in reviewing survey papers, our observations reveal that the majority of the articles reviewed are featured in the survey papers focused on higher education, possibly because there is enhanced data accessibility facilitated by the implementation of Learning Management Systems (LMS) in higher education institutions. Additionally, conducting scientific experiments is often more feasible within the higher education context. Broadly speaking, the survey papers categorized EDM methods into three groups: prediction, clustering, and regression. These papers also outlined the ML or the DM algorithms associated with each method category.

TABLE I. Previous literature reviews

Ref#	Publication year	Covered period	Paper reviewed	Research Questions
[6]	2015	2002–2015	39	1- What are the important attributes used in predicting students’ performance? 2- What are the prediction methods used for students performance?
[7]	2018	2010-2017	497	1- How is performance defined? What types of metrics are used for describing student performance? 2- What are the features used for predicting performance? 3-What methods are used for predicting performance? 4- Which feature and method combinations are used to predict which types of student performance? 5- What is the quality of the work on predicting student performance?
[8]	2020	2010-2020	67	1- Learning Outcomes Prediction. How is student academic performance measured using learning outcomes? 2-Academic Performance Prediction Approaches. What intelligent models and techniques are devised to forecast student academic performance using learning outcomes? 3-Academic Performance Predictors. What dominant predictors of student performance using learning outcomes are reported?

[9]	2020	2015-2019	120	<p>1- What are the methods and algorithms used to evaluate academic performance?</p> <p>2- What is the performance per method?</p> <p>3-: What are the features used to evaluate academic performance?</p> <p>4- To what extent were the results of the research used for decision making, according to the authors?</p>
[10]	2021	2009-2021	78	<p>1- What type of problems exist in the literature for Student Performance Prediction?</p> <p>2- What solutions are proposed to address these problems?</p> <p>3- What is the overall research productivity in this field?</p>
[11]	2021	2010-2020	90	<p>1- What are the most algorithms of supervised machine learning that widely used in prediction?</p>
[12]	2021	2010 to 2020	48	<p>1- What are the most commonly used methods in students' performance prediction?</p> <p>2- Which one is the most suitable method for students' performance prediction among the commonly used methods?</p>
[13]	2021	2015-2019	30	<p>1- What attributes are used for the prediction of student's academic performance?</p> <p>2- What are the key machine learning approaches used for the prediction?</p> <p>3- What are the accuracies of the existing models used in the prediction?</p>
[14]	2021	2016-2021	80	<p>1- Procedure of establishing prediction model. What procedure do researchers follow to establish the students' performance prediction model? What are the main steps?</p> <p>2- The EDM methods used in different steps. What EDM methods do researchers use in different steps in the procedure of building students' performance prediction model?</p> <p>3- Main challenges of previous studies. What are the challenges for previous studies in this field?</p>

[15]	2021	2015-2020	12	<p>1- What are the student performance measures to be predicted?</p> <p>2- What are the predictors used to train an explainable model?</p> <p>3- What are the explainable machine learning methods used to predict students' performance?</p> <p>4- What are the evaluation metrics used to assess the explainability of the models?</p> <p>5- What are the methods that meet both requirements of high accuracy and explainability?</p>
[16]	2022	2017 to 2022	18	None
[17]	2022	2016-2022	54	<p>1- What factors are determinants to predict academic performance?</p> <p>2- What methods and algorithms are applicable to predict academic performance?</p> <p>3- What are the goals and interests to predict academic performance?</p>
[18]	2022	2015 to 2022	35	<p>1- Identification and understanding of various researches done in the Educational sector for APP through a comprehensive and systematic review.</p> <p>2- Enumeration of various attributes used for APP.</p> <p>3- Analysis of ML techniques used in APP.</p> <p>4- Categorization of the combination of attributes and MI techniques to assess the accuracy of prediction.</p>
[19]	2022	2013-2021	33	<p>1- What are the primary purposes of the review studies investigated in this overview?</p> <p>2- What common input (predictor) and common output (target) variables do these review studies employ to predict SAP.</p> <p>3- What common educational data mining (EDM) techniques (or methods) and algorithms do they employ in predicting SAP?</p> <p>4- What algorithms are reported to have the highest prediction accuracy for SAP?</p>

				<p>5- What common EDM tools do these studies employ in predicting SAP?</p> <p>6- What are the key results of these review studies?</p>
[20]	2022	2017-2021	100	<p>1- Student performance factors: (A) What kind of data sources and factors can help us to determine high-risk students and (B) what factors fail to forecast the performance and behavior of students?</p> <p>2- Learning Analytics Approaches: What learning analytics approaches can be used to improve the student's performance and teaching practice?</p> <p>3- Machine Learning-based Approaches: What kind of algorithms are mostly used by researchers to develop predictive models in student retention with high accuracy?</p>
[21]	2022	2015-2020	40	<p>1- What is the most commonly used methodology for applying data mining?</p> <p>2- What attributes were considered for the prediction of academic performance?</p> <p>3- Which variable selection algorithms are most commonly used?</p> <p>4- What were the techniques used in the prediction and which had better results in their accuracy?</p> <p>5- Which tools are the most concurrent for the development and testing of the predictive model?</p> <p>6- What metrics are used to determine the effectiveness of prediction techniques?</p>
[22]	2023	2016-2022	84	<p>1- Which characteristics are frequently employed by academics when predicting student learning outcomes?</p> <p>2- Which ML methods are frequently employed by researchers when predicting student learning outcomes?</p> <p>3- Which algorithms or techniques are most effective for predicting student performance?</p>

Furthermore, the majority of reviews focused on technical aspects, providing a comprehensive account of the methods, techniques, and research objectives. Nevertheless, a comprehensive analysis of the procedures

used to establish students' performance prediction models was lacking, and there was no overview of the latest EDM methods utilized in steps other than model establishment. Due to the shortcomings in the mentioned surveys, researchers face challenges in identifying the essential processes involved in establishing prediction models for students' performance across different studies. Moreover, gaining an understanding of commonly used EDM methods and their efficacy in various stages of constructing prediction models is difficult, hindering efforts to optimize these models further. Simultaneously, investigating and summarizing the primary challenges and future directions in this research field becomes a challenging task. Finally, one aspect that remains unexplored quantitatively is the linkage between research outcomes and educational policy. This connection can span from the micro level, involving individual classes, to the macro level, encompassing the broader educational system. This review aims to investigate whether educational administrators have utilized research findings, obtained through technical data mining, in decision-making processes.

METHODOLOGY

To explore the most recent techniques in the field EDM employed by researchers for forecasting student performance, this study conducts a thorough and structured analysis of significant research focusing on predictive models for student performance using EDM. Furthermore, the study aims to consolidate research that introduces interpretable models for forecasting student performance. To achieve this, a simplified systematic literature review approach is adopted, facilitating the identification, selection, and comprehensive assessment of pertinent scholarly works that revolve around the anticipation of students' academic performance using EDM. Following the outlined methodology and the literature survey, this study undertakes a comparative and evaluative investigation of EDM techniques utilized in key stages such as methodology and tools, attributes, outcomes, selection algorithms, techniques, and metrics. In particular, we delve into the subsequent research inquiries:

- 1) Which methods are commonly employed for the implementation EDM?
- 2) What are the data sources utilized for EDM?
- 3) What are software tools utilized for EDM
- 4) What are the objectives considered in students' performance prediction?
- 5) What are the attributes taken into account when predicting academic performance?
- 6) What prediction techniques were employed?
- 7) What metrics are employed to assess the efficacy of prediction techniques?

In this survey, a search queries was performed in online databases including Google Scholar, IEEE Xplore, and Science Direct. The search criteria encompassed terms such as ("prediction) AND ("student performance") AND ("Machine Learning" OR "Data Mining"). Our focus was directed towards methods rooted in Machine Learning (ML) or Data Mining (DM) for the purpose of predicting student performance. We confined our attention to studies published in the English language between 2015 and 2022, Table II shows the number of papers that satisfied the search string. With such a huge number of articles in this filed the following inclusion and exclusion criteria are used for further filtering:

A. Inclusion

- Factors that influence academic performance.

- Prediction through data mining.
- Academic performance prediction model.

B. Exclusion

- University students as the only study population.
- Published outside the search period range.

TABLE II. Articles Data Sources

Databases	URL	Results
IEEE	https://ieeexplore.ieee.org/	167
Science Direct	https://www.sciencedirect.com/	227
Springer	https://link.springer.com/	25

Out of the 1951 papers surveyed in the mentioned databases, the initial phase identified 443 papers. Following this, the titles and abstracts of the articles underwent filtering using inclusion criteria, resulting in the exclusion of 239 articles. A total of 140 articles remained for in-depth analysis of the full text. After thoroughly examining all texts and rigorously applying inclusion and exclusion criteria, a final count of 74 articles was determined, forming the basis for the comprehensive literature review.

RESULTS AND FINDINGS

Based on the chosen literature, seven pertinent research aspects were identified to address the posed questions: methodologies, data source, software tools, objectives of the studies, attributes/factors, ML techniques, and metrics. This section will present the related work and outcomes in alignment with the six queries outlined in the methodology section.

Which methods are commonly employed for the implementation EDM?

Educational Data Mining (EDM) relies on various methodologies and frameworks, such as Knowledge Discovery in Databases (KDD), Cross-Industry Standard Process for Data Mining (CRISP-DM), Sample, Explore, Modify, Model, and Assess (SEMMA), and Team Data Science Process (TDSP), to effectively analyze educational data and derive valuable insights for improved educational outcomes. KDD’s comprehensive approach, covering data selection, preprocessing, transformation, and interpretation, suits the diverse challenges in education. CRISP-DM, known for its structured phases, finds application from understanding the educational context to model deployment. SEMMA, introduced by Statistical Analysis System (SAS) Institute, focuses on Sample, Explore, Modify, Model, and Assess, providing a systematic framework for preprocessing and modeling. Microsoft’s TDSP offers collaborative and scalable methodologies tailored to data science projects, including education. These methodologies empower educators, researchers, and institutions to harness data-driven insights for personalized learning and continuous improvement. Choosing the most suitable methodology depends on the specific context and goals of each EDM project. By analyzing selected papers, figure (2) reveals a prevalent adoption of the KDD methodology, though some researchers explicitly state the use of CRISP-DM, while SEMMA and TDSP methodologies see limited utilization. The preference for KDD in EDM stem its adaptability, holistic nature, and iterative process, aligning well with the dynamic aspects of event data management tasks. The benefits of employing KDD methodology can be outlined as follows:

- 1) KDD provides a systematic and structured process for analyzing data. This structured approach helps researchers and educators to organize their efforts efficiently and effectively.
- 2) KDD methodology offers techniques to handle highly complex diverse types of information contained in Educational data and extract meaningful patterns and insights.
- 3) KDD emphasizes feature selection, which is crucial in educational data mining to identify relevant variables that influence student learning outcomes.
- 4) KDD methodologies often employ scalable algorithms and techniques that can handle data of varying sizes, making them suitable for analyzing educational data at different levels of granularity.
- 5) KDD is an iterative process, allowing researchers to refine their analysis based on feedback and domain knowledge.

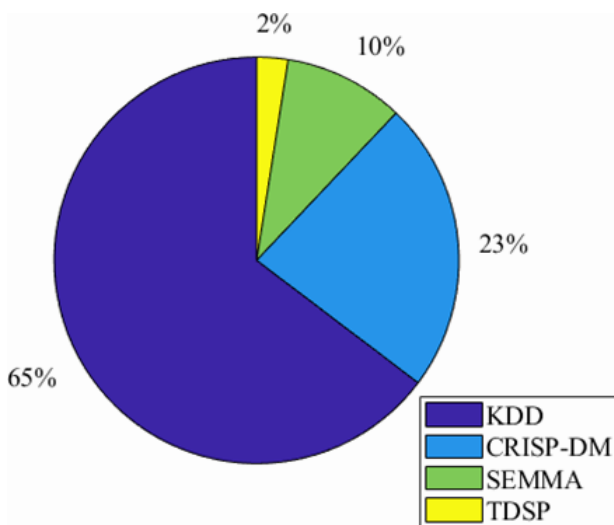


Fig. 2. Data Mining Methodologies Distribution.

A. What are the data sources utilized for EDM?

Educational Data Mining (EDM) relies on a diverse array of data sources to glean meaningful insights into the learning process, student performance, and educational system dynamics. One primary source is Learning Management Systems (LMS), comprehensive platforms that capture data on student interactions, engagement patterns, and academic progress. These systems record student logins, time spent on various learning materials, and assessment results, providing a rich source for analyzing learning behaviors. Additionally, Student Information Systems (SIS) contribute crucial demographic information, enrollment history, and academic records, enabling a holistic view of student profiles. Assessment data, including standardized test scores, quizzes, and assignments, serves as a valuable source for evaluating student performance and identifying areas for improvement. Regarding surveys and questionnaires, Tools like Google Forms, SurveyMonkey, and Qualtrics allow you to create and distribute surveys to gather information about student preferences, attitudes, and experiences. Finally, online discussion platforms like Piazza, Edmodo, discussion forums and social interactions within online learning platforms, discussion forums, and collaborative tools can be used to collect data on student interactions during discussions and group activities. In addition to aforementioned data sources, the utilization of publicly available datasets from the internet is a common practice in EDM. Researchers and analysts often leverage these datasets to study educational patterns, assess learning outcomes, and derive insights. The availability of such datasets contributes to the advancement of EDM by offering a diverse range of educational data for analysis and

model development. Researchers can explore publicly accessible educational databases to conduct studies, design models, and enhance understanding in the field of education through data-driven approaches. Depending on the objectives of the research and resources available for the researchers, one or more of the listed approaches can be adopted. Table III provides the list of the commonly utilized data sources. Based on the findings, SIS stands out as the most widely adopted data source for EDM development. Among the 74 studies examined, 33 opted for SIS when undertaking their data mining initiatives, and 18 utilized dataset available through the internet.

TABLE III. Common Data Sources used in EDM

Data Source	Article Reference
SIS	[23], [24], [26], [27], [28], [29], [32], [33], [37], [38], [40], [43], [45], [46], [48], [49], [55], [57], [61], [62], [63], [64], [65], [68], [69], [72], [73], [83], [88], [89], [92], [93], [94]
LMS	[34], [47], [52], [56], [71], [74], [81], [85], [91]
Surveys and Questionnaires	[25], [30], [35], [41], [44], [53], [57], [72], [86], [90]
Public Dataset (Internet)	[31], [36], [42], [51], [54], [57], [59], [60], [67], [75], [76], [80], [82], [84], [87], [91], [95], [96]
Others	[39], [66], [77], [78]

Figure (3), shows the percentage distribution of each data sources. The widespread adoption of SIS in educational institutions is driven by their ability to centralize, integrate, and provide access to student data, making them invaluable tools for educational data mining initiatives aimed at improving student outcomes and enhancing teaching and learning processes. It should be mentioned here that in recent years internet-based datasets are found to offer a wealth of opportunities for educational data mining research, providing rich, diverse, and large-scale data that can inform the development of more effective teaching and learning strategies, support evidence-based decision-making in education, and advance our understanding of learning processes in digital environments.

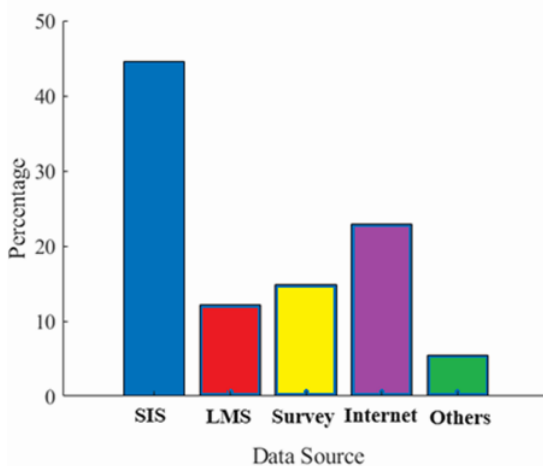


Fig. 3 Data Sources distribution.

B. What are the software tools utilized in EDM?

Since DM in general is a process of discovering patterns, trends, and valuable insights from large datasets, special software tools are required for this purposes. These software tools used in data mining serve various

functions and provide a wide range of capabilities to help analysts and data scientists extract valuable insights from large datasets. They offer a range of features and capabilities, such as Data Preprocessing, Data Transformation, Exploratory Data Analysis, Classification and Prediction and many other functionalities. The choice of tool depends on the specific requirements of a data mining project, as well as personal preferences and the organization’s budget considerations. Numerous tools are available for data mining, and each offers various data management capabilities. Table IV highlights some of the commonly employed tools for constructing predictive academic performance models. Based on the findings, Weka stands out as the most widely adopted tool for data mining development. Out of the 74 studies analyzed, Weka was chosen in 25 instances, while RapidMiner was utilized in 8 studies for their data mining endeavors. Both Weka and RapidMiner are open-source platforms tailored for data mining and machine learning tasks. Weka contains a comprehensive array of algorithms encompassing classification, regression, clustering, association rules, and feature selection, all applicable to educational datasets. Conversely, RapidMiner offers a user-friendly graphical interface facilitating data preprocessing, modeling, evaluation, and visualization, alongside an array of machine learning and statistical algorithms suited for EDM tasks. On the other hand, the combination of ease of learning, abundance of libraries, open-source nature, community support, interdisciplinary applications, scalability, and industry relevance makes Python a popular choice in the field of EDM. It provides a powerful and accessible platform for students and educators to explore, analyze, and visualize data in educational settings. Finally, as it has been noticed that among the studies, 21 are grouped under “others.” In these cases, the authors have omitted mentioning the software platform or programming language employed for their analysis. Furthermore, some of these studies have devised their own systems without specifying the software platform utilized in their development.

TABLE IV. Common Data Sources used in EDM

Software	Article Reference
Python Programming	[36], [42], [44], [50], [51], [52], [56], [59], [60], [67], [71], [80], [85], [87], [90], [95], [96]
Weka Software	[23], [25], [27], [28], [29], [30], [33], [35], [38], [39], [43], [45], [49], [55], [63], [69], [72], [73], [74], [77], [84], [91], [93], [94]
R Programming	[31], [54], [61], [79], [82]
RapidMiner Software System	[26], [32], [40], [41], [46], [48], [78]
Others	[24], [34], [37], [47], [53], [57], [58], [62], [64], [65], [66], [68], [70], [75], [76], [81], [83], [86], [88], [89], [92]

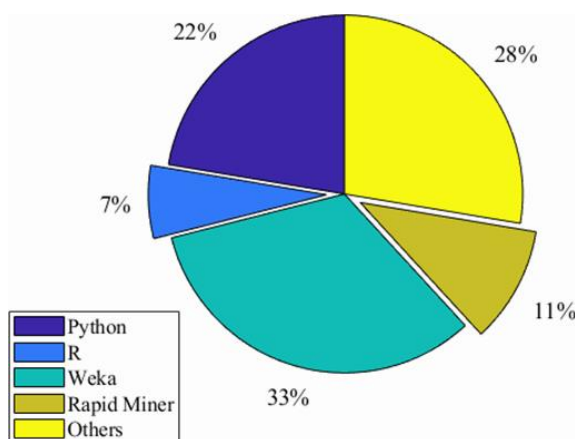


Fig. 4. Percentage Distribution of Utilized Software Tools.

C. What are the objectives considered in students’ performance prediction?

Measuring students’ performance can be done through various outcomes and assessment methods. The choice of outcomes depends on the educational goals, subject matter, and grade level. When using machine learning to predict or assess students’ performance, various outcomes can be considered. These predicted outcomes can provide valuable information for modeling and predicting student performance. Machine learning models can use wide range of outcomes to give certain indication about a student’s performance, such as the likelihood of passing a course or achieving a certain grade. It’s important to note that the choice of outcome and the success of the prediction model depend on the availability and quality of data, as well as the specific goals of the assessment. Based on the collected reference, table V lists the most widely predicted outcomes in students’ academic performance are: grades predictions, early warning systems, personalized learning, graduation and retention rates, adaptive learning and college admissions.

Figure (5) shows the main objectives or targets that has been considered in the reviewed papers. The prediction of students’ grades as can be noticed from the figure have attracted around 50% of the studies since it is a crucial objective in educational data mining because it supports efforts to improve student outcomes, personalize learning experiences, allocate resources effectively, prevent dropout, ensure institutional accountability, and empower students to succeed academically. By leveraging predictive analytics, educators and educational institutions can make data-driven decisions that enhance teaching and learning processes and ultimately contribute to the success of all students. The second widely objective is the personalized and adaptive learning which scored around 40% of the reviewed papers. Analyzing personalized learning in educational data mining helps educators optimize instruction, support individual student needs, foster academic success, and improve learning outcomes for all students. By leveraging data-driven insights, educators can create more engaging, effective, and equitable learning experiences that meet the diverse needs of learners in today’s classrooms.

TABLE V. Main Objectives of the Studies

Objectives	Article Reference
Predicting Grades	[25], [26], [27], [28], [30], [31], [32], [33], [36], [38], [41], [42], [43], [44], [45], [46], [49], [53], [54], [55], [59], [62], [63], [64], [65], [66], [67], [68], [70], [73], [74], [79], [80], [83], [93], [96]
Early Warning Systems	[23], [24], [28], [29], [30], [36], [39], [50], [51], [54], [56], [69], [71], [72], [73], [76], [77], [81], [82], [88], [94]
Personalized and Adaptive Learning	[24], [26], [28], [29], [33], [35], [37], [40], [42], [44], [47], [48], [52], [58], [60], [71], [73], [74], [76], [78], [84], [85], [87], [90], [91], [92], [95]
Graduation and Retention	[34], [39], [41], [44], [45], [56], [92], [94]
College Admissions	[23], [39], [64], [68], [74], [86], [89], [93]

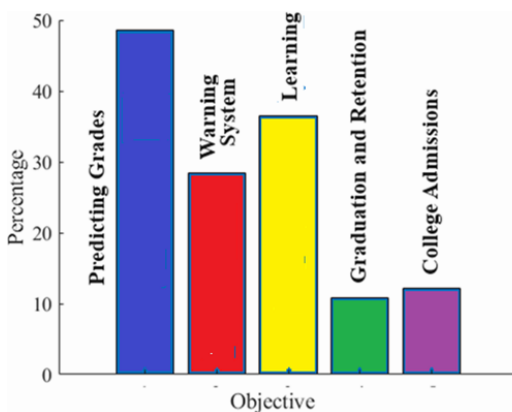


Fig. 5. Distribution Objectives in the Primary Studies.

D. What are the attributes taken into account when predicting academic performance?

Forecasting student performance is an intricate undertaking that entails scrutinizing an array of characteristics and elements. These attributes are harnessed to construct models capable of approximating a student’s probable academic achievement. In response to this inquiry, we successfully recognized that there are enormous number of attributes used in the literature. Through thorough examination of the existing literature, we have identified that these attributes can be classified into ten main categories as listed in table VI. It’s important to recognize that each category contains several attributes, and in some instances, certain attributes have been assigned to different categories in the surveyed studies. Furthermore, most of the surveyed studies investigated a wide range of these attributes.

TABLE VI. Ain Attributes.

Category	Attributes	Article Reference
Personal	Age, Gender, knowledge and background, Physical health, Employment Status, Personal interests..... etc.	[23], [24], [25], [28], [29], [30], [31], [33], [34], [35], [36], [37], [41], [42], [45], [46], [47], [49], [50], [51], [54], [55], [57], [63], [69], [70], [71], [73], [74], [75], [77], [79], [80], [84], [87], [88], [89], [91], [92], [93], [94], [95], [96]
Academic performance	GPA, CPGA, Major, prior courses grades, present quizzes, assignments, exams marks, Major, enrollment year, attendance, discussion number of fails, Motivation,... etc.	[24], [25], [26], [27], [28], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [45], [46], [47], [48], [49], [52], [54], [55], [56], [57], [58], [60], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [74], [75], [76], [77], [79], [80], [81], [83], [84], [85], [87], [88], [89], [90], [91], [92], [93], [94], [96]
Studying Environments	Type of the institute, class size, Access to books, research materials, and technology, Availability of study aids, Adequate internet connectivity for online research and learning, ... etc.	[23], [24], [25], [29], [30], [35], [36], [42], [44], [47], [70], [73], [75], [77], [79], [80], [84], [88], [91], [95]
Studying style	Academic participation, Time Spent on daily study, Group study, Individual study, Utilization of library, utilization of the internet in the study,... etc.	[29], [30], [35], [36], [41], [42], [44], [47], [50], [51], [53], [54], [57], [63], [70], [71], [74], [75], [79], [80], [81], [85], [86], [90], [96]
Prior Academic Achievement	High school location and type, Admission score, Admission type, Entrance exam score, pre-university subject grades... etc.	[25], [26], [29], [30], [32], [33], [35], [39], [47], [50], [51], [56], [57], [61], [63], [64], [68], [69], [75], [77], [80], [81], [88], [93], [95]
Family	Family size, Father and mother occupations, Father and mother qualification, Family income, level of parental involvement etc.	[23], [24], [25], [30], [25], [36], [42], [50], [51], [54], [56], [57], [60], [65], [69], [70], [89], [93], [96]

Social	Internet Access, Relationship, friends, social networking, non-academic activities, Social Media, ... etc.	[25], [30], [35], [36], [38], [41], [42], [50], [51], [53], [56], [57], [67], [70], [75], [77], [79], [80], [85], [87], [93], [96]
Demographics	Race, Nationality, Place of Birth, Marital status, Hometown, Cultural background	[23], [25], [29], [30], [33], [35], [36], [38], [45], [47], [50], [51], [54], [56], [57], [63], [66], [67], [68], [69], [70], [71], [74], [75], [79], [80], [84], [85], [87], [91], [93], [95], [96]
Behavioral	Personality, Self-discipline, Initiative, Perseverance, Ethics, Positive peer influence, Collaboration, ... etc.	[29], [30], [35], [37], [41], [42], [53], [57], [66], [75], [80], [85], [86]
Financial	Scholarship, financial aid, self-spaceship, has income, Family Financial Situation, Cost of Living, type of accommodation... etc.	[24], [25], [26], [30], [31], [35], [36], [37], [42], [51], [52], [54], [58], [70], [74], [76], [80], [81], [88], [92], [96], [97]

As it can be noticed from table VI and figure 6 that academic factors are most frequently used factors since they are crucial components that significantly influence the prediction of a student’s performance. These factors pertain directly to a student’s educational experiences and achievements within an academic context. Within this context, the characteristics that saw the highest frequency of utilization were associated with the prior academic achievement, field of study (major) and teaching environments. It was found that the CGPA stands out as the predominant attribute due to its concrete numerical value, which directly links it to students’ performance, explaining its inclusion in more than 82.5% of the research studies. Furthermore, the total credits also emerges as one of the attributes displaying the strongest correlation with the outcome variable.

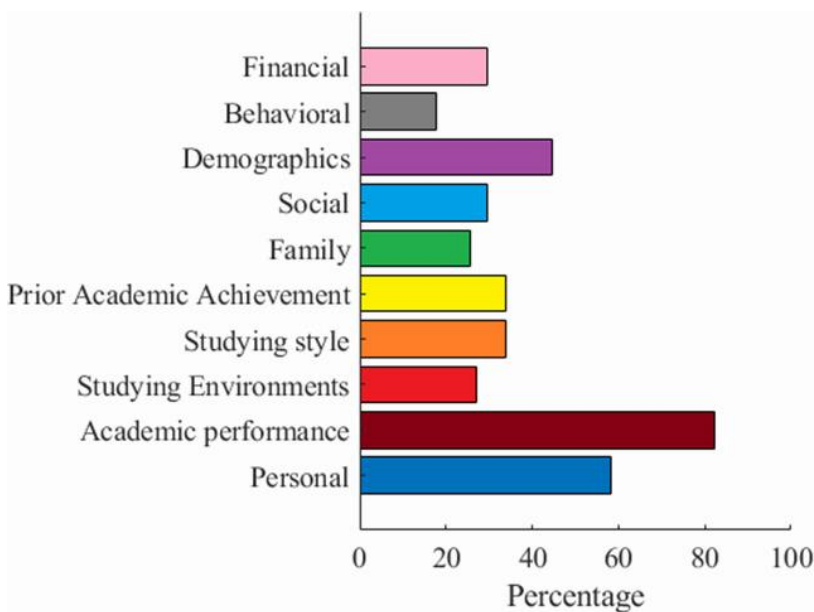


Fig. 6. Distribution of Investigated Attributes Categories.

The second influential set of attributes are related to personal factors, in which 58% of the studies considered these attributes in their analysis. These factors can vary from one individual to another and can have both direct and indirect effects on a student’s academic performance. The significance of each attribute’s contribution is also emphasized, clearly indicating that Gender and Age play a crucial role in predicting academic performance. Gender is a prevalent variable across numerous studies, as it exerts a substantial influence on academic achievement. This is evident in a large number of studies, where gender is

highlighted as one of the most influential attributes in predictive models. According to these studies, being female increases the likelihood of academic success, potentially attributed to the learning strategies employed by women during their academic journeys. Coupled with their responsible approach to learning, these factors contribute to a more effective teaching and learning process for female students.

The third influential set of attributes are the demographic factors in which 45% studies utilized some of these factors. Demographic factors are characteristics related to an individual's identity and background that can influence student performance prediction. These factors provide important context for understanding a student's educational experiences and outcomes. It's essential to note that while demographic factors can play a role in performance prediction, they do not determine a student's destiny, and individual outcomes can vary widely.

Studying style and prior academic achievement attributes have been investigated in around 34% of the studies. The impact of studying style on students' academic performance can be multifaceted and plays a crucial role in shaping students' academic performance by influencing learning efficiency, time management, motivation, engagement, retention, recall, study environment preferences, and adaptability. Recognizing and leveraging individual studying styles can empower students to maximize their learning potential and achieve academic success. On the other hand, prior academic achievement, which refers to a student's previous performance in academic endeavors, can have a significant impact on their current academic performance. Prior academic achievement significantly impacts student academic performance by providing a strong foundation of knowledge and skills, fostering confidence and self-efficacy, shaping study habits and learning strategies, opening access to opportunities, influencing teacher expectations and support, and shaping peer interactions and the academic environment. Recognizing the importance of prior academic achievement can inform educational practices aimed at promoting student success and maximizing learning outcomes.

In the fifth tier of significance, both social and financial factors have each accounted for approximately 30% of utilization in the studies surveyed. Social factors such as peer influence, parental support, socioeconomic status, school climate, cultural expectations, peer pressure, and social support networks profoundly impact students' academic performance. Recognizing and addressing the influence of social factors is essential for promoting equitable access to education and fostering positive learning environments that support all students in achieving academic success. In contrast, financial factors significantly influence students' academic performance by shaping their access to resources, quality of education, educational opportunities, basic needs, academic support services, college affordability, and psychological well-being. Addressing financial disparities and ensuring equitable access to educational resources and opportunities are essential for promoting academic success and reducing educational inequities.

The studying environment plays a crucial role in shaping students' academic performance where 27% of the studies considered these factors in their investigations. In general, studying environments significantly influence students' academic performance by shaping their physical, social, technological, and cultural experiences. Creating supportive, inclusive, and conducive studying environments that prioritize comfort, minimize distractions, foster peer support, leverage technology effectively, and embrace cultural diversity can enhance students' motivation, engagement, and learning outcomes. At the same level of the studying environment, family's attributes have been investigated in approximately 26%. Family attributes, including parental involvement, education level, socioeconomic status, family structure, parenting style, cultural values, and parental support and expectations, collectively influence students' academic performance. Recognizing the importance of family factors in shaping educational outcomes can inform efforts to support students' academic success and promote equity in education.

Finally the lowest set of attributes that have been considered in the surveyed studies is the behavioral attributes, only 17.5% of the papers have investigated these attributes on the students' academic

performance. In general, behavioral attributes such as motivation, work ethic, time management skills, organization, self-regulation, responsibility, engagement, resilience, and adaptability significantly influence students’ academic performance. Cultivating positive behavioral traits and providing support for developing essential skills are essential for promoting academic success and empowering students to reach their full potential.

The selection of variables, also known as feature selection, can have a significant impact on the prediction of students’ academic performance. Selecting relevant variables that are strongly correlated with academic performance can improve the accuracy of predictive models. Including these relevant features in predictive models allows the model to capture meaningful patterns and relationships that contribute to academic performance. Moreover, Feature selection helps in reducing the dimensionality of the dataset by eliminating irrelevant or redundant variables. High-dimensional datasets with many features can lead to overfitting, decreased model interpretability, and increased computational complexity. By selecting a subset of the most informative features, feature selection simplifies the model while retaining important predictive information, leading to more efficient and interpretable models. Furthermore, feature selection helps in improving the generalization of predictive models by reducing the risk of overfitting. Overfitting occurs when a model learns noise or irrelevant patterns from the training data, leading to poor performance on unseen data. Likewise, this stage enhances the interpretability of predictive models by focusing on a subset of relevant features that are easier to understand and interpret. Models with fewer features are more transparent and intuitive, allowing stakeholders, such as educators, policymakers, and administrators, to gain insights into the factors influencing academic performance and make informed decisions based on the model’s predictions. Finally, feature selection improves the computational efficiency of predictive modeling by reducing the computational burden associated with training and evaluating models on high-dimensional datasets. Selecting a subset of relevant features reduces the number of calculations required for model training, evaluation, and inference, leading to faster model training times and lower computational costs.

TABLE VII. Feature Selection Methodologies.

Feature Selection Methodology	Article Reference
Correlation-Based Attribute	[30], [35], [42], [46], [49], [59], [71], [80], [90], [91], [93], [94]
Gain Ratio-Based Attribute	[29], [39], [43], [46], [49], [73], [91]
Chi – squared	[39], [45], [53], [66]
Minimum Redundancy Maximum Relevance	[31], [43], [54], [95]
Dimensionality Reduction	[42], [63], [70], [88], [96]
Cross-Validation	[59], [67], [70], [80], [84]
Manual	[33], [50], [61]
Others	[36], [56], [64], [65], [75], [76], [89]

As been noticed approximately 50% of the surveyed papers performed feature selection process. Different approaches but most of these methods are based on Filter Methods which includes statistical measures in rank correlation coefficient as in the Correlation-Based Attribute and Chi-squared, Information Gain and Gain Ratio to rank the features based on their ability to reduce uncertainty about the target variable as in the Gain Ratio-Based Attribute. Table VII list the identified feature selection methodologies in the surveyed papers.

E. What prediction techniques were employed?

Predicting students’ performance using EDM and ML involves the analysis of educational data to uncover

patterns and make predictions about student outcomes. Three prediction categories were identified: Classification Algorithms, Clustering Algorithms, and Regression Algorithms. Classification techniques are widely used to predict whether a student will pass or fail a course or achieve a specific performance level. Common classification algorithms include decision trees, logistic regression, support vector machines, and random forests. Clustering techniques are used to group students with similar learning behaviors or profiles. These groups can then be analyzed to understand common characteristics that influence performance. K-Means clustering and hierarchical clustering are often applied in EDM. Finally, Discovery techniques are used to predict and model continuous variables related to student performance. Table VIII, presents many of these techniques and the obtained accuracy when applied to this field.

TABLE VIII. General Machine Learning Prediction Categories

Approach	Article Reference
Prediction and Classification	[23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [35], [36], [37], [38],[39], [40], [41], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [58], [59], [60], [62], [63], [65], [65], [66], [67], [70], [71], [73], [74], [76], [77], [78], [79], [80], [81], [83], [84], [85], [86], [87], [88], [89], [90], [91], [93], [94], [95], [96]
Clustering	[50], [72], [78], [82], [88]
Knowledge Discovery	[34], [42], [56], [57], [61], [69], [75], [92]

As it can be inferred from table VIII that the EDM research often focuses on student performance classification since it provides predictive analysis through the understanding patterns in student data, educators can predict future performance. Also classifying students based on their academic performance, educators can provide personalized learning plans, adaptive resources, and targeted interventions. Furthermore, the classification can identify which concepts students struggle with the most, enabling educators to refine and improve teaching materials and instructional strategies can inform decisions about curriculum design, teacher training, and educational standards. Finally, classifying students based on their academic performance, educators can implement targeted interventions such as extra tutoring, mentoring programs, or counseling services.

In the majority of articles reviewed, multiple algorithms were employed. To organize the researchers' approaches, we categorized the algorithms into various approaches, including Decision Trees, Naive Bayes, K-Nearest Neighbor, Logistic Regression, Neural Networks, Support Vector Machine, Deep Learning and Other Methods. Table IX and figure 7 provide an overview of the frequencies of the measurement tools utilized. Notably, accuracy emerged as the most commonly employed evaluation tool.

TABLE IX. Commonly Used Machine Learning Techniques.

Technique	Article Reference
Decision Tree	[23], [24], [25], [27], [28], [29], [31], [33], [35], [36], [37], [38], [39], [43], [44], [45], [46], [47], [48], [49], [51], [52], [57], [59], [62], [63], [64], [67], [69], [71], [73], [74], [76], [78], [79], [80], [84], [85], [87], [89], [90], [91], [92], [93], [95]
K-Nearest Neighbor	[24], [43], [46], [57], [59], [62], [63], [64], [67], [69], [70], [71], [73], [76], [78], [80], [83], [87], [89], [95], [96]
Naive Bayes	[24], [28], [29], [30], [31], [33], [38], [44], [46], [47], [50], [51], [55], [57], [67], [69], [70], [73], [74], [76], [78], [83], [85], [86], [91], [92], [93], [94], [96]
Neural Network	[24], [29], [33], [38], [43], [44], [51], [56], [57], [60], [62], [70], [71], [72], [73], [80], [91], [95]

Random Forest	[28], [38], [40], [42], [43], [44], [51], [56], [67], [69], [71], [76], [83], [85], [86], [88], [89], [90], [91], [92], [93], [94], [95]
Rule-Based	[24], [26], [92]
Support Vector Machine	[29], [36], [41], [43], [44], [51], [53], [57], [59], [60], [61], [67], [69], [70], [71], [76], [80], [83], [85], [86], [87], [88], [89], [90], [95]
Deep Learning	[58], [65], [67]
Bayesian Network	[24], [30]
Linear Regression	[31], [32], [43], [54], [61], [62], [69], [70], [71], [79], [89], [95], [96]
Logistic Regression	[46], [51], [56], [60], [83], [86]
K-Means	[50], [72], [82]
Others	[33], [34], [42], [43], [52], [56], [59], [66], [70], [71], [75], [76], [77], [80], [81], [87], [88], [90], [92], [94], [95], [96]

A notable prevalence of Decision Tree algorithms was observed, nearly 62% utilized this technique in their analysis. This high occurrence can be attributed to both the method’s widespread adoption and the abundance of algorithms accessible in commonly used tools like Python and WEKA. Decision Tree algorithms constituted the predominant choice due to their demonstrated high accuracy. Within this category, various algorithms such as C4.5, ID3, CART, Random Trees, among others, offer researchers a diverse array of approaches to select from, often within the same article. Bayesian algorithms, specifically Naive Bayes, were also widely used because of the simplicity, efficiency, scalability, and competitive classification performance of Naive Bayes make it a popular choice in educational data mining, where the focus is often on analyzing large and complex datasets to support decision-making and improve educational outcomes. The Support Vector Machine achieved a score of around 34%. This result aligns with the perspective of numerous experts in the field who consider this technique advantageous due to its reputed high accuracy, resilience against overfitting, adaptable kernel functions, suitability for small sample sizes, interpretability, capacity to handle irrelevant features, and broad applicability, rendering Support Vector Machines a favored choice in educational data mining endeavors

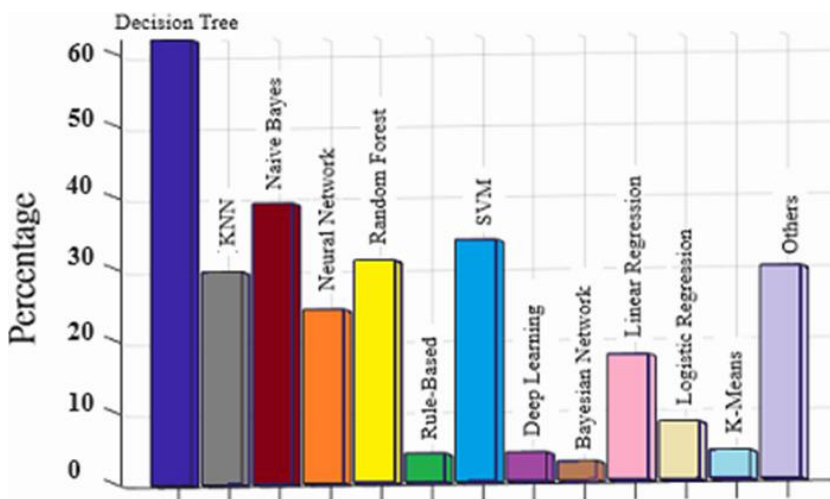


Fig. 7. Distribution of Prediction Techniques

F. What metrics are employed to assess the efficacy of prediction techniques?

When using ML techniques to predict students’ academic performance, several performance measures are commonly employed to assess the accuracy and effectiveness of predictive models. These metrics are

essential for evaluating the performance of machine learning techniques because they provide quantitative assessment, serve as benchmarks for comparison, offer insights into model behavior, identify strengths and weaknesses, guide model selection and tuning, support decision-making, and facilitate communication of results. In Table X, you can observe the metrics employed for assessing the effectiveness of prediction models. These metrics offer insights into whether the model under development aligns closely with real-world data. Various metrics have been employed by authors across different studies, with Accuracy (59%), Precision (47.3%), Recall (50%) and F-Measure (39.1%) being the most widely favored metric. This is primarily because it facilitates the evaluation of predictive model quality. Additionally, other frequently utilized metrics such as Mean Absolute Error, Area under the Curve, ROC Area and others are also used in a number of investigations.

DISCUSSION AND CONCLUSIONS

EDM stands out as a sophisticated and highly efficient approach for analyzing large volumes of educational data. Among the various research areas within EDM, predicting students' academic performance emerges as a predominant focus. It can be stated that huge amount of the studies in the EDM domain are dedicated to this particular aspect. Our survey delves into the analysis of 74 studies conducted over the past eight years, which have employed EDM techniques to forecast students' performance. This survey aims to furnish researchers in this field with a comprehensive understanding and practical guidance. We systematically examine the process of constructing predictive models for students' performance using EDM methodology. This involves a thorough discussion and comparison of the EDM techniques employed in the key stages of previous studies. Furthermore, seven research questions have been investigated individually and underscore the principal findings for each one.

TABLE X. Commonly Used Metrics.

Technique	Article Reference
Accuracy	[24], [26], [31], [33], [36], [38], [39], [40], [41], [42], [44], [47], [48], [50], [53], [56], [57], [59], [62], [63], [67], [69], [70], [71], [72], [73], [75], [76], [77], [78], [80], [82], [85], [86], [87], [89], [90], [91], [92], [93], [94], [96]
Precision	[24], [26],[27], [30], [31], [35], [40], [43], [47], [48], [51], [55], [56], [57], [59], [62], [63], [69], [70], [73], [74], [75], [78], [80], [82], [83], [86], [87], [88], [89], [90], [91], [92], [95], [96]
Recall	[24], [25], [26], [27], [28], [30], [31], [35], [40], [43], [47], [48], [49], [51], [55], [56], [57], [59], [62], [63], [70], [73], [74], [78], [80], [82], [83], [86], [87], [88], [89], [90], [91], [92], [93], [95], [96]
F-Measure	[24], [28], [29], [31], [40], [43], [48], [49], [51], [55], [56], [57], [59], [63], [69], [70], [73], [74], [76], [80], [82], [83], [87], [88], [92], [93], [94], [95], [96]
Mean Absolute Error	[24], [52], [54], [65], [66], [90]
ROC Area	[27], [51], [59], [77], [80], [92], [93]
Sensitivity	[41], [50], [53]
Specificity	[41], [50]
Area Under the Curve	[46], [52], [63], [70], [77], [83], [87], [91], [96]
Root Mean Squared Error	[52], [53], [54], [65], [69], [90], [91]
Confusion Matrix	[23], [28], [62], [63], [65], [69]

Accuracy	[24], [26], [31], [33], [36], [38], [39], [40], [41], [42], [44], [47], [48], [50], [53], [56], [57], [59], [62], [63], [67], [69], [70], [71], [72], [73], [75], [76], [77], [78], [80], [82], [85], [86], [87], [89], [90], [91], [92], [93], [94], [96]
other	[61], [65], [66], [67], [69], [73], [77], [81], [87], [90], [92], [93]

Throughout the development of prediction models, researchers have identified four critical issues and two general criticisms with various methodologies from EDM and ML, the outline of the critical concluding issues are summarized here:

1) Since the KDD methodology is highly adopted in EDM application, SIS and LMS would most certainly serve as valuable data sources when employing KDD as an EDM methodology for several reasons. Firstly, these systems contain comprehensive and diverse datasets encompassing student demographics, academic performance, attendance records, and learning activities. This rich pool of data provides a holistic view of student behavior and performance, facilitating in-depth analysis and pattern recognition. Secondly, SIS and LMS data are typically well-structured and standardized, this will simplify the preprocessing and feature extraction stages of the KDD process. Additionally, the availability of historical data in these systems enables longitudinal studies and trend analysis, enhancing the predictive capabilities of the EDM models. Lastly, the widespread adoption of SIS and LMS across educational institutions ensures the accessibility and scalability of data, allowing researchers to conduct large-scale analyses and derive meaningful insights to support decision-making and improve educational outcomes.

2) Software and programming tools play a crucial role in predicting student academic performance through EDM for several reasons. Firstly, these tools facilitate data preprocessing and transformation, allowing researchers to clean, normalize, and aggregate heterogeneous data from various sources such as Student Information Systems (SIS) and Learning Management Systems (LMS). Effective preprocessing is essential for ensuring data quality and consistency, which directly impacts the accuracy and reliability of predictive models. Secondly, software and programming tools provide a platform for implementing and testing different machine learning algorithms and data mining techniques. Researchers can experiment with a wide range of algorithms to identify the most suitable approaches for predicting student academic performance. In conclusion, these tools offer flexibility and scalability, enabling researchers to analyze large datasets efficiently and iteratively refine their models based on performance metrics and domain knowledge. Furthermore, software and programming tools support the interpretation and visualization of model results, allowing researchers to gain insights into the underlying factors influencing student performance. Visualizations such as feature importance plots, decision trees, and confusion matrices help elucidate the relationships between input variables and academic outcomes, aiding educators in identifying at-risk students and implementing targeted interventions. These tools enable researchers to harness the power of data to develop accurate and actionable insights that can inform educational practices and improve student outcomes.

3) Specifying an objective is essential for guiding the EDM process, ensuring relevance and alignment with stakeholders' needs, facilitating evaluation and validation, optimizing resource allocation, and promoting ethical conduct in predicting student academic performance. It should be noted that the features or attributes used in predicting student academic performance through EDM vary based on the specific objective, contextual considerations, data availability, model complexity, and stakeholder input. Researchers must carefully consider these factors when selecting features to ensure that prediction models are effective, interpretable, and actionable in addressing the intended objectives. It should be emphasized that in the prediction of students' academic performance a high volume of data is essential for developing accurate, reliable, and generalizable predictive models in educational data mining. It allows researchers to capture the complexity of student behavior, represent diverse features, train and evaluate models effectively, address data imbalances, and handle variability inherent in educational datasets. Another aspect in this critical issue

is related to data preprocessing. Data preprocessing is essential when predicting student academic performance using educational data mining because it improves data quality, enhances model performance, ensures algorithm compatibility, and enhances interpretability. These preprocessing steps are crucial for obtaining accurate and reliable predictions that can inform educational decision-making and support student success. The final aspect in this category is related to feature selection. Feature selection is essential when predicting students' academic performance using educational data mining because it helps reduce dimensionality, improve model performance, enhance interpretability, optimize resource efficiency, and enhance robustness and generalization. These benefits ultimately contribute to the development of more accurate, efficient, and actionable predictive models that support student success in education.

4) Selecting the appropriate machine learning method for forecasting students' academic achievements through EDM involves considering numerous factors, such as the particular problem at hand, the attributes of the data, the need for interpretability, and the considerations for generalization. Given the intricate and diverse nature of educational data and the multifaceted task of predicting academic performance, it's difficult for a specific machine learning technique to universally outperform others across all scenarios and settings. The optimal choice of technique typically hinges on the specific attributes of the dataset, the intricacies of the prediction task, and the interpretive demands of educational stakeholders. Rather than striving for a single superior technique, researchers and practitioners often explore a spectrum of methods and strategies, capitalizing on their respective strengths and trade-offs to construct effective predictive models tailored to unique educational contexts and objectives. It is found that the majority of researchers employ supervised classification algorithms from the realm of machine learning to construct prediction models. Among these algorithms, commonly utilized ones include Decision Trees, Naive Bayes, Neural Networks, Random Forests, Support Vector Machines, K-Nearest Neighbors, and others. As been stated that there was a significant prevalence of Decision Tree algorithms, with approximately 62% of researchers utilizing this technique in their analyses. This high frequency can be attributed to both the method's widespread adoption and the availability of numerous algorithms within commonly used tools, furthermore, Decision Tree algorithms were the primary choice due to their proven high accuracy. Finally, as it has been stated Metrics play a crucial role in evaluating the effectiveness and performance of machine learning models. These metrics provide quantitative measures that help assess how well a model is performing relative to the desired outcomes. Importantly, metrics provide insights into various aspects of a model's performance, including its accuracy, precision, recall, F1-score, and more, depending on the specific task and objectives. These metrics allow practitioners to gauge the model's ability to correctly classify instances, identify false positives and false negatives, and strike a balance between different evaluation criteria. Moreover, metrics enable comparisons between different models or variations of the same model, aiding in the selection of the most suitable approach for a particular problem. They also help in identifying potential areas for improvement and fine-tuning the model parameters or features.

5) While notable advancements have been made by researchers in this domain, there remain, many limitations and large improvement spaces still exist. Firstly, researchers predominantly employ EDM methods directly derived from the fields of DM or ML, without tailoring or optimizing these techniques to suit the unique characteristics of educational data and the prediction of students' performance. Secondly, scant attention is paid to the interpretability of prediction models, resulting in a lack of clarity in the prediction process. Consequently, educators face challenges in discerning the factors significantly influencing students' academic performance, rendering the prediction results dubious. Thirdly, there exists a necessity for enhancing the quality of data utilized for training prediction models. Specifically, researchers ought to gather a more extensive dataset comprising historical academic data from diverse student populations, while also standardizing dataset parameters.

Overall, this review has identified a rich collection of analysis methods as well as a predominant focus on tertiary education. A limited application of data mining methods has been found to support educational

policy-making and institutional decision-making. We believe that the development of research aimed at their own application in the daily teaching process but also in support of decision making at the level of educational policy, should be an alternative. Internal feedback with successful examples of using different algorithms and techniques in the long run, without practical application, can lead to the scientific field withering away.

REFERENCES

1. Trolan, T.L., "Student Engagement in Higher Education: Conceptualizations, Measurement, and Research", In: Perna, L.W. (eds) Higher Education: Handbook of Theory and Research. Higher Education: Handbook of Theory and Research, vol 39. Springer, Cham, 2024.
2. Ünal, "Data Mining for Student Performance Prediction in Education", Data Mining-Methods, Applications and Systems", IntechOpen, Jan. 20, 2021.
3. Alejandro Peña-Ayala, "Educational Data Mining Applications and Trends", Springer Cham, 2014. Namoun, Abdallah, and Abdullah Alshamqiti, "Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review", Applied Sciences, vol. 11, no. 1, 237.
4. I. Papadogiannis, V. Pouloupoulos, and M. Wallace, "A Critical Review of Data Mining for Education: What has been done, what has been learnt and what remains to be seen", International Journal of Educational Research Review, vol. 5, no. 4, pp. 353–372, 2020.
5. Albreiki, Balqis, Nazar Zaki, and Hany Alashwal, "A Systematic Literature Review of Student Performance Prediction Using Machine Learning Techniques", Education Sciences, vol. 11, no. 9, 552, 2021.
6. Khalaf, Alaa S., Jasim Mohammed Dahr, Ihab Ahmed Najim, Mohammed B. M. Kamel, Ali Salah Hashim, Wid Aqeel Awadh and Aqeel Majeed Humadi. "Supervised Learning Algorithms in Educational Data Mining: A Systematic Review", Southeast Europe Journal of Soft Computing, vol. 10, no. 1, pp.55-70, 2021.
7. Lynn N. D.Emanuel A. W. R., "Using Data Mining Techniques to Predict Students' Performance. a Review", IOP Conf. Ser.: Materials Science and Engineering, vol. 1096, pp. 1-9, 2021.
8. Y. Baashar, G. Alkaws, N. Ali, H. Alhussian and H. T. Bahbouh, "Predicting student's performance using machine learning methods: A systematic literature review," 2021 International Conference on Computer & Information Sciences (ICCOINS), Kuching, Malaysia, pp. 357-362, 2021.
9. Xiao W, Ji P, Hu J., "A survey on educational data mining methods used for predicting students' performance", Engineering Reports, vol. 4, no. 5, pp. 1-23, 2022.
10. R. Alamri and B. Alharbi, "Explainable Student Performance Prediction Models: A Systematic Review," in IEEE Access, vol. 9, pp. 33132-33143, 2022.
11. Rimpay, A. Dhankhar and K. Solanki, "Educational Data Mining tools and Techniques used for Prediction of Student's Performance: A Study," 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, pp. 1-5, 2022.
12. De-La-Cruz, P., Rojas-Coaquira, R., Vega-Huerta, H., Pérez-Quintanilla, J., and Lagos-Barzola, M., "A Systematic Review Regarding the Prediction of Academic Performance", Journal of Computer Science, vol. 8, no. 12, pp. 1219–1231, 2022.
13. P. Chakrapani and C. D., "Academic Performance Prediction Using Machine Learning: A Comprehensive & Systematic Review," 2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC), Chennai, India, pp. 335-340, 2022.
14. Chaka, C., "Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: an overview of reviews", Journal of E-Learning and Knowledge Society, vol. 18, no. 2, pp. 58-69, 2022.
15. D. A. Shafiq, M. Marjani, R. A. A. Habeeb and D. Asirvatham, "Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review," in IEEE Access, vol. 10, pp.

72480-72503, 2022.

16. Alfredo Daza, Carlos Guerra, Noemí Cervera and Erwin Burgos, “Predicting Academic Performance through Data Mining: A Systematic Literature”, TEM Journal. vol. 11, issue 2, pp. 939-949, 2022.
17. Iddrisu Issah, Obed Appiah, Peter Appiahene and Fuseini Inusah, “A systematic review of the literature on machine learning application of determining the attributes influencing academic performance”, Decision Analytics Journal, vol. 7, pp. 1-11, 2023.
18. S. M. Merchan Rubiano and J. A. Duarte Garcia, “Formulation of a predictive model for academic performance based on students’ academic and demographic data,” 2015 IEEE Frontiers in Education Conference (FIE), El Paso, TX, USA, pp. 1-7, 2015.
19. P. Amornsinlaphachai, “Efficiency of data mining models to predict academic performance and a cooperative learning model”, 2016 8th International Conference on Knowledge and Smart Technology (KST), Chiang Mai, Thailand, pp. 66-71, 2016.
20. Amjad Abu Saa, “Educational Data Mining & Students’ Performance Prediction”, International Journal of Advanced Computer Science and Applications, vol. 7, no. 5, pp. 212-220, 2016.
21. Yasmeen Altujjar, Wejdan Altamimi, Isra Al-Turaiki and Muna Al-Razgan, “Predicting Critical Courses Affecting Students Performance: A Case Study”, Procedia Computer Science, vol. 82, pp. 65 – 71, 2016.
22. Mashael A. Al-Barrak and Muna Al-Razgan , “Predicting students final GPA using decision trees A Case Study”, International Journal of Information and Education Technology, vol. 6, no. 7, pp. 528-533, July 2016.
23. T. Mahboob, S. Irfan and A. Karamat, “A machine learning approach for student assessment in E-learning using Quinlan’s C4.5, Naive Bayes and Random Forest algorithms,” 2016 19th International Multi-Topic Conference (INMIC), Islamabad, Pakistan, pp. 1-8, 2016.
24. Evandro B. Costa, Balduino Fonseca, Marcelo Almeida Santana, Fabrísia Ferreira de Araújo and Joilson Rego, “Evaluating the effectiveness of educational data mining techniques for early prediction of students’ academic failure in introductory programming courses”, Computers in Human Behavior, vol. 73, pp. 247-256, 2017.
25. Khalaf, Alaa and Humadi, Aqeel and Awadh, Wid Akeel and Hashim and Ali Salah, “Students’ Success Prediction Based on Bayes Algorithms”, International Journal of Computer Applications, vol. 178, no. 7, pp. 6-12, 2017.
26. Murat Pojon, Using Machine Learning to Predict Student Performance, Master Thesis, University of Tampere, 2017.
27. A. Anzer, H. A. Tabaza and J. Ali, “Predicting Academic Performance of Students in UAE Using Data Mining Techniques” 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), Paris, France, pp. 179-183, 2018.
28. Bergen Karabulut and Şeyma Cihan, “Educational Data Mining Application for Increasing Quality in Engineering Education”, The Online Journal of Quality in Higher Education, vol. 5, issue 3, pp. 7-12, July 2018.
29. Sumyea Helal, Jiuyong Li, Lin Liu and Esmaeil Ebrahimie, Shane Dawson, Duncan J. Murray, “Identifying key factors of student academic performance by subgroup discovery”, Int J Data Sci Anal, vol. 7, pp. 227–245, 2019.
30. Khalaf, Alaa and Hashim, Ali Salah and Awadh, Wid Akeel, “Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis”, International Journal of Interactive Multimedia and Artificial Intelligence vol. 5.2, pp. 26-31, 2018.
31. X. Ma and Z. Zhou, “Student pass rates prediction using optimized support vector machine and decision tree,” 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, pp. 209-215, 2018.
32. G. -H. Wang, J. Zhang and G. -S. Fu, “Predicting Student Behaviors and Performance in Online Learning Using Decision Tree” 2018 Seventh International Conference of Educational Innovation through Technology (EITT), Auckland, New Zealand, pp. 214-219, 2018.
33. C. -C. Kiu, “Data Mining Analysis on Student’s Academic Performance through Exploration

- of Student's Background and Social Activities" 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), Subang Jaya, Malaysia, pp. 1-5, 2018.
34. Ace C. Lagman and Rossana T. Adao, "Embedding Machine Learning Algorithm Models in Decision Support System in Predicting Student Academic Performance Using Enrollment and Admission Data" Proceedings of 2018 the 8th International Workshop on Computer Science and Engineering, Bangkok, pp. 298-302, 2018.
 35. N. Ketui, W. Wisomka and K. Homjun, "Using Classification Data Mining Techniques for Students Performance Prediction" 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON), Nan, Thailand, pp. 359-363, 2019.
 36. I. Burman and S. Som, "Predicting Students Academic Performance Using Support Vector Machine" 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, pp. 756-759, 2019.
 37. H. K. R. Toppireddy, B. Saini and P. S. Hada, "Academic Enhancement System using EDM Approach," 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India, pp. 1-10, 2019.
 38. B. Al Breiki, N. Zaki and E. A. Mohamed, "Using Educational Data Mining Techniques to Predict Student Performance," 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA), Ras Al Khaimah, United Arab Emirates, pp. 1-5, 2019.
 39. M. R. Rimadana, S. S. Kusumawardani, P. I. Santosa and M. S. F. Erwianda, "Predicting Student Academic Performance using Machine Learning and Time Management Skill Data," 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, pp. 511-515, 2019.
 40. Y. T. Samuel, J. J. Hutapea and B. Jonathan, "Predicting the Timeliness of Student Graduation Using Decision Tree C4.5 Algorithm in Universitas Advent Indonesia," 2019 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, pp. 276-280, 2019.
 41. S. Shrestha and M. Pokharel, "Machine Learning algorithm in educational data," 2019 Artificial Intelligence for Transforming Business and Society (AITB), Kathmandu, Nepal, pp. 1-11, 2019.
 42. E. N. Azizah, U. Pujianto, E. Nugraha and Darusalam, "Comparative performance between C4.5 and Naive Bayes classifiers in predicting student academic performance in a Virtual Learning Environment" 2018 4th International Conference on Education and Technology (ICET), Malang, Indonesia, pp. 18-22, 2018,
 43. N. Ketui, W. Wisomka and K. Homjun, "Using Classification Data Mining Techniques for Students Performance Prediction," 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON), Nan, Thailand, pp. 359-363, 2019.
 44. I. Khan, A. Al Sadiri, A. R. Ahmad and N. Jabeur, "Tracking Student Performance in Introductory Programming by Means of Machine Learning," 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), Muscat, Oman, pp. 1-6, 2019.
 45. Y. Widyaningsih, N. Fitriani and D. Sarwinda, "A Semi-Supervised Learning Approach for Predicting Student's Performance: First-Year Students Case Study," 2019 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, pp. 291-295, 2019.
 46. Deepti Aggarwal, Sonu Mittal and Vikram Bali, "Prediction Model for Classifying Students Based on Performance using Machine Learning Techniques", International Journal of Recent Technology and Engineering (IJRTE), vol. 8, issue 2S7, , pp. 496-503, 2019.
 47. Tsiakmaki, Maria, Georgios Kostopoulos, Sotiris Kotsiantis, and Omiros Ragos, "Implementing AutoML in Educational Data Mining for Prediction Tasks" Applied Sciences 10, no. 1: 90, 2020.
 48. Farhatul Janan and Sourav Kumar Ghosh, "Prediction of Student's Performance Using Support Vector Machine Classifier", Proceedings of the 11th Annual International Conference on Industrial

- Engineering and Operations Management Singapore, pp. 7078-7088, 2021.
49. El Aissaoui, O., El Alami El Madani, Y., Oughdir, L., Dakkak, A. and El Alloui, Y., “A Multiple Linear Regression-Based Approach to Predict Student Performance”, In: Ezziyyani, M. (eds) *Advanced Intelligent Systems for Sustainable Development (AI2SD’2019)*. AI2SD 2019. *Advances in Intelligent Systems and Computing*, Springer, Cham, vol. 1102, pp. 9-23, 2019.
 50. Baz, Abdullah, Fatima Alshareef, Ebtihal Alshareef, Hosam Alhakami and Tahani Alsubait. “Predicting Students’ Academic Performance Using Naïve Bayes.” *IJCSNS International Journal of Computer Science and Network Security*, vol. 20 no. 4, pp. 182-190, 2020.
 51. Almutairi, S., Shaiba, H. and Bezbradica, M., “Predicting Students’ Academic Performance and Main Behavioral Features Using Data Mining Techniques”, In: Alfaries, A., Mengash, H., Yasar, A., Shakshuki, E. (eds) *Advances in Data Science, Cyber Security and IT Applications. ICC 2019. Communications in Computer and Information Science*, vol 1097, Springer, Cham.
 52. 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,
 53. Li, Shuping and AU – Liu, Taotang, “Performance Prediction for Higher Education Students Using Deep Learning”, *Complexity, Hindawi*, vol. 2021, 2021.
 54. J. Chowdhery, A. Jasmin, A. Jaiswal and J. A. A. Jothi, “Automatic Student Performance Prediction System Using Data Mining Techniques,” *2021 International Conference on Computing and Communications Applications and Technologies (I3CAT)*, Ipswich, United Kingdom, pp. 57-65, 2021.
 55. Priya S, Ankit T and Divyansh D, “Student Performance Prediction Using Machine Learning”, in *Advances in Parallel Computing Technologies and Applications*, vol. 40, pp. 167–174, 2021.
 56. Michael McDonald and Yulei Pang, “Comparing the Performance on the MI and SAT/PSAT for the Purpose of Monitoring Student Achievement”, *International Journal of Information and Education Technology*, vol. 11, no. 12, pp. 590-596, 2021.
 57. N. Rachburee, P. Sunantapot, D. Ounjit, P. Panklom, P. Porking and W. Punlumjeak, “A Major Recommendation System in Educational Mining,” *2021 1st International Conference On Cyber Management And Engineering (CyMaEn)*, Hammamet, Tunisia, pp. 1-5, 2021.
 58. J. Cui, Y. Zhang, R. An, Y. Yun, H. Dai and X. Shang, “Identifying Key Features in Student Grade Prediction”, *2021 IEEE International Conference on Progress in Informatics and Computing (PIC)*, Shanghai, China, pp. 519-523, 2021.
 59. Hussain, S., Khan and M.Q., “Student-Performulator: Predicting Students’ Academic Performance at Secondary and Intermediate Level Using Machine Learning”, *Ann. Data. Sci.* vol. 10, pp. 637–655, 2023.
 60. A. Nabil, M. Seyam and A. Abou-Elfetouh, “Prediction of Students’ Academic Performance Based on Courses’ Grades Using Deep Neural Networks,” *IEEE Access*, vol. 9, pp. 140731-140746, 2021.
 61. T. Fang, S. Huang, Y. Zhou and H. Zhang, “Multi-model Stacking Ensemble Learning for Student Achievement Prediction” *2021 12th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP)*, Xi’an, China, pp. 136-140, 2021.
 62. Opeyemi Ojajuni, Foluso Ayeni, Olagunju Akodu, Femi Ekanoye, Samson Adewole, Timothy Ayo, Sanjay Misra and Victor Mbarika, “Predicting Student Academic Performance Using Machine Learning” In: Gervasi, O., et al. *Computational Science and Its Applications – ICCSA 2021. ICCSA 2021. Lecture Notes in Computer Science*, vol 12957. Springer, Cham., 2021.
 63. Shatha Ghareeb, Abir Hussain, Wasiq Khan, Dhiya Al-Jumeily, Thar Baker and Rawaa Al-Jumeily, “Dataset of student level prediction in UAE”, *Data in Brief*, vol. 35, pp. 1-6, 2021.
 64. Palacios, Carlos A., José A. Reyes-Suárez, Lorena A. Bearzotti, Víctor Leiva, and Carolina Marchant, “Knowledge Discovery for Higher Education Student Retention Based on Data Mining: Machine Learning Algorithms and Case Study in Chile” *Entropy*, vol. 23, no. 4: 485, 2021.
 65. R. Ayienda, R. Rimiru and W. Cheruiyot, “Predicting Students Academic Performance using a Hybrid of Machine Learning Algorithms,” *2021 IEEE AFRICON*, Arusha, Tanzania, United Republic of, pp. 1- 6, 2021.

66. Trakunphutthirak, R., and Lee, V. C. S, “Application of Educational Data Mining Approach for Student Academic Performance Prediction Using Progressive Temporal Data”, *Journal of Educational Computing Research*, vol. 60, no. 3, pp. 742-776, 2022.
67. Khaledun Nahar, Boishakhe Islam Shova, Tahmina Ria, Humayara Binte Rashid and A. H. M. Saiful Islam, “Mining educational data to predict students’ performance”, *Education and Information Technologies*, vol. 26, pp. 6051–6067, 2021.
68. Ijaz Khan, Abdul Rahim Ahmad, Nafaa Jabeur and Mohammed Najah Mahdi, “An artificial intelligence approach to monitor student performance and devise preventive measures”, *Smart Learn. Environ.*, vol. 8, no. 17, pp. 1-18, 2021.
69. Agung Triayudi and Wahyu Oktri Widyarto, “Comparison J48 and Naïve Bayes Methods in Educational Analysis”, *Virtual Conference on Engineering, Science and Technology (ViCEST)*, pp. 1-5. 2020.
70. Wen Xiao, Ping Ji, Juan Hu, “RnkHEU: A Hybrid Feature Selection Method for Predicting Students’ Performance”, *Scientific Programming*, vol. 2021, 2021.
71. E. Evangelista, “A Hybrid Machine Learning Framework for Predicting Students’ Performance in Virtual Learning Environment”, *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 24, pp. 255–272, 2021.
72. Flores, Vaneza, Stella Heras, and Vicente Julian. “Comparison of Predictive Models with Balanced Classes Using the SMOTE Method for the Forecast of Student Dropout in Higher Education”, *Electronics*, vol. 11, no. 3: 457, 2022.
73. M. B. Issa, O. Darwish, D. H. Allah, F. Shatnawi, D. Darweesh and Y. M. Tashtoush, “Analysis of Jordanian University Students Problems Using Data Mining System,” 2022 13th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, pp. 220-225, 2022.
74. Yağcı, M, “Educational data mining: prediction of students’ academic performance using machine learning algorithms”, *Smart Learn. Environ.* Vol. 9, no. 11,
75. A. Garg, N. B. Garg, P. Ghosh, A. Bansal, U. K. Lilhore and S. Simaiya, “A Machine Learning-based Automatic Model to Predicting Performance of Students”, 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET), Bhopal, India, pp. 1-5, 2022.
76. Lassaad K. Smirani, Hanaa A. Yamani, Leila Jamel Menzli and Jihane A. Boulahia, “Using Ensemble Learning Algorithms to Predict Student Failure and Enabling Customized Educational Paths”, *Scientific Programming*, vol. 2022, 2022.
77. G. Fan and Y. Chen, “Analysis and Prediction of Students’ Academic Performance Based on Educational Data Mining,” *IEEE Access*, vol. 10, pp. 19558-19571, 2022.
78. L. E. Contreras Bravo, N. Nieves-Pimiento, and K. Gonzalez-Guerrero, “Prediction of University-Level Academic Performance through Machine Learning Mechanisms and Supervised Methods”, *Ing.*, vol. 28, no. 1, p. e19514, 2022.
79. Sultan Juma Sultan Al Alawi, Jastini Mohd Jamil and Izwan Nizal Mohd Shaharane, “Predicting Student Performance Using Data Mining Approach: A Case Study in Oman”, *MSEA*, vol. 71, no. 4, pp. 1389–1398, 2022.
80. Y. Joshi, K. Mallibhat and V. M., “Students’ Performance Prediction Using Multimodal Machine Learning,” 2022 IEEE IFEES World Engineering Education Forum – Global Engineering Deans Council (WEEF-GEDC), Cape Town, South Africa, pp. 1-5, 2022.
81. Wang D, Lian D, Xing Y, Dong S, Sun X and Yu J, “Analysis and Prediction of Influencing Factors of College Student Achievement Based on Machine Learning”, *Front. Psychol.*, vol. 13, pp. 1-13, 2022.
82. Amal Shaker and Abdelmoniem Helmy, “Predicting Students Academic Performance Using Machine Learning Techniques”, *EasyChair Preprint* no. 9194, 2022.
83. E. Alhazmi and A. Sheneamer, “Early Predicting of Students Performance in Higher Education,” *IEEE Access*, vol. 11, pp. 27579-27589, 2023.
84. Abideen, Zain ul, Tehseen Mazhar, Abdul Razzaq, Inayatul Haq, Inam Ullah, Hisham Alasmay, and Heba G. Mohamed, “Analysis of Enrollment Criteria in Secondary Schools Using Machine Learning and Data Mining Approach”, *Electronics*, vol. 12, no. 3: 694, 2023.

85. Alruwais, Nuha, and Mohammed Zakariah, “Evaluating Student Knowledge Assessment Using Machine Learning Techniques”, *Sustainability*, vol. 5, no. 7: 6229, 2023.
86. Padmalaya Nayak, Sk. Vaheed, Surbhi Gupta and Neeraj Mohan, “Predicting students’ academic performance by mining the educational data through machine learning-based classification model”, *Educ Inf Technol*, vol. 28, pp. 14611–14637, 2023.
87. Sánchez, Aurora, Cristian Vidal-Silva, Gabriela Mancilla, Miguel Tupac-Yupanqui, and José M. Rubio, “Sustainable e-Learning by Data Mining—Successful Results in a Chilean University”, *Sustainability* vol. 15, no. 2: 895, 2023.
88. Alghamdi, Amnah Saeed, and Atta Rahman, “Data Mining Approach to Predict Success of Secondary School Students: A Saudi Arabian Case Study”, *Education Sciences*, vol.13, no. 3: 293, 2023.
89. Meghji AF, Mahoto NA, Asiri Y, Alshahrani H, Sulaiman A and Shaikh A, “Early detection of student degree-level academic performance using educational data mining”, *PeerJ Computer Science*, vol. 9, 2023.
90. Orlando Iparraguirre-Villanueva, Carmen Torres-Ceclén, Andrés Epifanía-Huerta, Gloria Castro-Leon, Melquiades Melgarejo-Graciano, Joselyn Zapata-Paulini and Michael Cabanillas-Carbonell, “Comparison of Predictive Machine Learning Models to Predict the Level of Adaptability of Students in Online Education”, *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 14, no. 4, 2023.
91. Xinning Wang, Yuben Zhao, Chong Li and Peng Ren, “ProbSAP: A comprehensive and high-performance system for student academic performance prediction”, *Pattern Recognition*, vol. 137, 2023.
92. Charrois, Theresa L., “Systematic reviews: what do you need to know to get started?”, *The Canadian journal of hospital pharmacy*, vol. 68, no. 2, pp. 144-148, 2015.
93. Lesley Uttley, Daniel S. Quintana, Paul Montgomery, Christopher Carroll, Matthew J. Page, Louise Falzon, Anthea Sutton and David Moher, “The problems with systematic reviews: a living systematic review”, *Journal of Clinical Epidemiology*, vol 156, pp. 30-41, 2023.
94. Shahiri, Amirah Mohamed, Wahidah Husain and Nur’Aini Abdul Rashid, “A Review on Predicting Student’s Performance Using Data Mining Techniques.” *Procedia Computer Science*, vol. 72, pp. 414-422, 2015.
95. Hellas, Arto, Petri Ihantola, Andrew Petersen, Vangel V. Ajanovski, Mirela Gutica, Timo Hynninen, Antti Knutas, Juho Leinonen, Christopher H. Messom and Soohyun Nam Liao, “Predicting academic performance: a systematic literature review”, *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, pp. 175–199, 2018.
96. Xinning Wang, Yuben Zhao, Chong Li and Peng Ren, “ProbSAP: A comprehensive and high-performance system for student academic performance prediction”, *Pattern Recognition*, vol. 137, 2023.