# Leveraging Machine Learning to Predict Academic Success: A Literature Review and Bibliometric Trend Analysis

### Abstract

Numerous review studies have explored machine learning (ML)’s role in predicting academic success. In this paper, we perform three complementary analyses: (1) a systematic review of 33 review articles drawn from Web of Science (WOS), SCOPUS, and Taylor & Francis to identify prevailing methodologies and gaps; (2) a qualitative synthesis of the 10 most cited WOS articles to highlight influential algorithms and interpretability techniques; and (3) a bibliometric analysis of 285 WOS articles to chart publication trends, country-level productivity, and international collaboration networks. Our structured findings reveal an increasing focus on ensemble methods and explainable artificial intelligence (AI; e.g., SHAP, LIME), sustained growth in publication volume (2019–2024), and a leading output from China alongside strong collaboration hubs in Australia, India, and Europe. Persistent challenges include inconsistent terminology, data heterogeneity, and limited real-world deployment of ML models.

### Keywords

Data Mining; Machine Learning; Academic Success; Educational Analysis; Systematic Literature Review; Bibliographical Analysis

## Introduction

In today’s complex world, predicting academic success has become a key focus in education research (Guanin-Fajardo et al., 2024) forming a key research theme in learning analytics (Jovanović et al., 2021). As the demand for skilled professionals grows, understanding what helps students succeed is more important than ever. This has led to a surge in studies aiming to predict academic outcomes, such as graduation rates and the likelihood of student dropouts, or grade point average itself.

Student academic success is a key indicator of educational quality indicating students’ behaviours aiming at accomplishing academic goals with persistence in course and degree completion (Crisp et al., 2015; Eather et al., 2015). The wide range of factors at educational, psychological, environmental, and social levels influence academic achievements (Tinto, 1975). York et al., (2015) define academic success as a sum of six core components: academic achievement, satisfaction, skill acquisition, persistence, learning objectives, and career success. Numerous success factors are often reduced to grade point average alone (Alyahyan & Düştegör, 2020). This overlaps with what students perceive as academic success, i.e. a combination of outcomes including grades and more holistic outcomes of personal development and achievements (Lynam et al., 2024). Academic success varies significantly among individuals (Schillereff et al., 2023), and is shaped by a combination of personal factors (Acosta-Gonzaga, 2023) such as self-regulation, motivation, and self-esteem; environmental factors (Edgerton & McKechnie, 2023) like socioeconomic status, school environment, and peer support; and lifestyle factors (Kassaw & Demareva, 2023) including dietary habits, sleep, and stress management. Predictive models of student academic performance are increasingly being used to identify at-risk students and provide timely interventions to improve retention rates and academic outcomes. These models leverage historical data, including demographic information, pre-academic performance, virtual learning environment (VLE) interactions, and assessment scores, to predict future academic performance (Umer et al., 2023).

### Predictive Models and Machine Learning in Education

Performance prediction has been studied as both a classification and regression problem, each with its own advantages and disadvantages (Alshanqiti & Namoun, 2020). A wide variety of statistical algorithms and techniques, including Machine Learning (ML), Statistical Analysis, and Deep Learning (DL), have been employed in the literature (Zhao et al., 2021). Predictive analytics in education, particularly within learning analytics, refers to the use of these techniques to analyze educational data and forecast student outcomes, such as academic performance or dropout risk. While ML focuses on pattern recognition and prediction using algorithms that learn from data, DL, a subset of ML, employs neural networks to model complex patterns. Statistical analysis, on the other hand, relies on traditional mathematical models to infer relationships. Comparing the performance of these algorithms is challenging due to the multi-sourced nature of the data (Zhao et al., 2021).

One of the most promising tools in this area of prediction is machine learning (Balaji et al., 2021; Jin, 2023). ML allows researchers to analyze large amounts of data to find patterns and make predictions, and in contrast to traditional statistical models, ML reduces bias, offers flexibility, and provides more robust models (Author 1, 2025; Hilbert et al., 2021). This technology can help identify students who might struggle, giving educators the chance to offer support early on. By using data like academic history and student behavior and, more recently, even online learning behaviors where open-access data is increasingly available (Liang et al., 2023), ML models can provide insights that help improve student outcomes. Jin (2023) employes four ML algorithms for predicting academic success by accurately identifying the varying impacts of student, parent, and school efforts on exam performance. These techniques, alongside statistical and deep learning methods, form the backbone of predictive analytics in education, enabling more personalized and proactive interventions.

### The Value of Bibliometric Analysis

Bibliometric analysis is a quantitative method used in systematic literature reviews to assess research productivity and impact (Lim & Kumar, 2024). Many seminal guides on this topic have been published, e.g., by Donthu et al., (2021); Lim & Kumar, (2024), focusing on performance metrics such as publication counts, citation analysis, authorship trends, keyword clustering, nomological networking, knowledge gaps, coverage of research, and so on. This technique leverages big data and software tools (e.g., *Gephi*, *VOSviewer*, or other already established statistics library tools such as *bibliometrics* package in *R*) to provide an objective overview of a research domain.

### Research Aims

This study aims to provide a comprehensive synthesis of ML applications in predicting academic success by conducting three complementary analyses. First, we systematically reviewed 33 existing review articles from WOS, SCOPUS, and Taylor & Francis databases to identify recurring trends, methods, and gaps in the literature. Second, we conducted a qualitative review of the 10 most cited WOS articles to gain deeper insights into influential algorithmic approaches, data types, and interpretability methods. Third, a bibliometric analysis of 285 articles exclusively from the WOS database was performed, enabling us to examine publication trends, co-authorship patterns, keyword evolution, and international research collaboration networks.

### Research Questions

The present review study is focused on research about ML methods for predicting academic success and therefore opens the following research questions:

RQ1: Based on our systematic review of 33 existing review articles (WOS, SCOPUS, Taylor & Francis), what recurring themes, methodological approaches, and research gaps characterize the literature on ML-based academic success prediction?

RQ2: From our qualitative synthesis of the 10 most cited WOS articles, which algorithmic strategies, data types, and interpretability techniques have been most influential in predicting academic success?

RQ3: Through bibliometric analysis of 285 WOS articles, how have publication patterns, country-level research productivity, and international collaboration networks evolved over time in this field?

## Methods

### Literature search

A preliminary analysis was conducted across *Web of Science (WOS), SCOPUS, and Taylor & Francis databases, as well as* the Digital Portal of the authors’ university(University registers). A comparison between the databases and registries (using the tool specified in *data analysis* section) yielded a high degree of overlap in studies related to the prediction of academic success using machine learning techniques. Articles from all selected databases and registries were accounted for in the review based on their relevance.

For the final report presented in this article, the initial search was first conducted in January 2025, and updated on February 10th. The search string applied was: *("machine learning" OR "ML") AND ("academic success" OR "academic performance" OR "student success") AND ("educational sciences" OR "education") AND (“prediction”)*. Search strings per database are presented in a Table 3 indicating a date of last search and number of hits.

|  |  |  |  |
| --- | --- | --- | --- |
| Database | Search string applied | Date of search | No. of results |
| Web of Science | ("machine learning" OR "ML") AND ("academic success" OR "academic performance" OR "student success") AND ("educational sciences" OR "education") AND (“prediction”)  (Used in all Fields) | 10. 2. 2025 | 292 |
| Scopus | ("machine learning" OR "ML") AND ("academic success" OR "academic performance" OR "student success") AND ("educational sciences" OR "education") AND (“prediction”) | 10. 2. 2025 | 7782 |
| Taylor & Francis | ("machine learning" OR "ML") AND ("academic success" OR "academic performance" OR "student success") AND ("educational sciences" OR "education") AND (“prediction”)  (Automatically transformed into [[All: "machine learning"] OR [All: "ml"]] AND [[All: "academic success"] OR [All: "academic performance"] OR [All: "student success"]] AND [[All: "educational sciences"] OR [All: "education"]] AND [All: "prediction"]) | 10. 2. 2025 | 484 |
| University register | ("machine learning" OR "ML") AND ("academic success" OR "academic performance" OR "student success") AND ("educational sciences" OR "education") AND (“prediction”) | 2. 8. 2025 | 1044 |

Table 3: Search strings, databases, date of search, number of results.

This study was conducted following the PRISMA guidelines to establish a rigorous, transparent, reproducible, and adaptable review (PRISMA-P Group et al., 2015; Page et al., 2022). The flowchart in Figure 1 indicates that the search resulted in 292 matches in WOS, 7782 matches in SCOPUS, Taylor and Francis resulted 484 matches, and the University registers 1044. For the bibliometric analysis, only filtered WOS records were included (n = 285).

Records removed *before screening*:

Records not in WOS (*n* =9310) – containing duplicates from other databases

Records published elsewhere than journals (*n* = 0).

Recors not in English (*n* = 0)

Records identified from\*:

Databases:

WOS (*n* = 292)

SCOPUS (*n* = 7782)

TAYLOR & FRANCIS (*n*=484)  
University registers (*n* = 1044)

**Identification**

Records excluded\*\*

Duplicate records (*n* = 1)

\*Records with final publication year in 2025: (*n* = 4)

Records deemed with Expression of Concern (*n* = 1)

Records deemed as Data Paper (*n* = 1)

Records screened based on title, abstract and key words

(*n* = 292)

**Screening**

Articles included in bibliographical analysis

(*n* = 285)

**Included**

Figure 1: PRISMA diagram.

\*Note: WoS differentiates between “publication date” and “final publication date”.

Inclusion and exclusion criteria were applied to ensure the relevance and quality of the data. Articles were included if they were written in English, published between 2019 and 2024, and classified as peer-reviewed articles (including 189 articles, 76 proceedings papers, 76 reviews, and 18 articles in early access). Both qualitative and quantitative studies were considered. Articles with expressions of concern or retractions were excluded to maintain the validity of the data. The years prior to 2019 and 2025 were omitted from the analysis for specific reasons: The results prior to 2019 were not included as the thematic focus during that year was still fairly new and in the early stages of progression (only 10 articles), making the data less relevant for current trends. Additionally, 2025 was excluded as the data from that year is still new and scarce, limiting its impact on the overall analysis.

### Bibliometric Data Analysis

The 285 WOS records were processed with custom Python scripts (Author 1, 2025) and the pyBibX library, which leverage WOS’s standardized fields—keywords, affiliations, citations—to construct co-authorship, keyword-co-occurrence, and citation networks. These scripts ensure reproducible extraction of publication years, citation counts, and geographic metadata.

## Results

### Existing review

Numerous review studies have already explored the application of ML in predicting academic success, as summarized in Table 2 in Appendix 7.2., where we analysed 33 review articles from WOS, SCOPUS and Taylor & Francis databases. These reviews systematically analyze the methodologies, datasets, and ML techniques employed across various educational contexts. The reviewed studies employed various systematic review methodologies, with the most common being the Systematic Literature Review (SLR), often following Kitchenham’s guidelines or the PRISMA framework, or other, and some not specifying the approach. Other methods such as meta-analysis, bibliometric analysis, scoping reviews, and systematic mapping are underrepresented. In terms of prediction levels, the majority of studies focused on degree-level outcomes, including graduation likelihood and dropout risk, course-level performance prediction, year-level academic success, and exam-level assessments. Other levels of prediction include computational thinking assessments, and citation prediction. ML techniques were predominant across all studies, with the most frequently applied methods being decision trees (DT), random forests (RF), support vector machines (SVM), artificial neural networks (ANN/MLP), logistic regression (LGR), and Naïve Bayes (NB), demonstrating their widespread use in predictive analytics for student performance. Hybrid models, ensemble methods, and clustering techniques were also commonly used. Notably, some studies explored deep learning (DNN, RNN, CNN, LSTM, Transformers) and reinforcement learning (Q-Learning, SARSA), with many studies reporting high predictive accuracy. However, recurring gaps such as limited real-world implementation, ethical concerns, and the underutilization of clustering and reinforcement learning techniques are noted in some studies. Despite the extensive research, there are only a limited number of studies employing bibliometric analysis, which could provide deeper insights into research trends, publication patterns, and thematic developments in this field.

### Most cited studies analysis

Subsequently, we conducted an extensive qualitative literature review of the 10 most cited articles from a systematic literature search. The selected articles, gathered in Appendix 7.1, published between 2021 and 2024 and cited between 88 and 143 times, collectively explore the application of ML and educational data mining (EDM) across various educational contexts, including higher education, online learning, and assessment strategies. The studies primarily focus on early performance prediction, student retention, and self-reported satisfaction in online education. Predominantly categorized under "Education & Educational Research; Computer Science, Information Systems; Engineering, Electrical & Electronic; Mathematics; Environmental Sciences; and Artificial Intelligence" in Web of Science, these articles have been published by major publishers such as Springer (four articles), and MDPI (three articles). The research employs various algorithms such as logistic regression, decision trees, random forests, support vector machines, XGBoost, and ensemble methods, with logistic regression and random forest often yielding the best performance, as demonstrated by Cankaya et al., (2024); Jang et al., (2022), while ensemble methods like GLMnet, as highlighted by (Bertolini et al., 2021) and wrapper-based feature selection techniques combined with classifiers like k-NN and SVM, as seen in (Abdelkader et al., 2022), are also widely employed to enhance predictive accuracy and interpretability, particularly through explainable AI (XAI; methods and techniques that make the outputs of machine learning models understandable to humans, enabling transparency and trust) techniques such as SHAP and LIME, enabling educators and policymakers to derive actionable insights. Key findings underscore the effectiveness of ensemble methods and the importance of feature selection in improving predictive accuracy. Conducted in diverse settings, including the USA, Spain, and international online platforms, these studies highlight the potential of ML to personalize educational interventions, mitigate attrition, and optimize instructional strategies, particularly in higher education and online learning environments. The list of the 10 most cited articles in the field, using the search string established in Methods section is gathered in Table 1 in Appendix 7.1.

### Bibliometric review

For word cloud analysis, the *CountVectorizer* module was used in collaboration with *Latent Dirichlet Allocation* (LDA) to extract the most frequently used words from keywords and abstracts. Common English stopwords (e.g., “the”, “and”, “is”) were removed to minimize noise, along with words appearing in more than 95% of the documents and those appearing in fewer than two documents.

Figure 2 presents annual word clouds from 2019 to 2024, capturing the evolving vocabulary in the domain of machine learning applications in education. While terms such as “machine learning” and “data” predictably dominate throughout, a year-by-year comparative analysis reveals significant thematic shifts and emergent research priorities. In 2019, the landscape is concentrated around broad concepts like prediction, learning, and machine learning, indicating an exploratory phase where educational applications were not yet deeply specialized. Terms like student and performance are relatively less emphasized, suggesting limited focus on individualized outcome prediction at this stage. By 2020, we observe the emergence of more targeted keywords such as student, performance, and classification. This suggests a pivot from general data-driven approaches to more specific applications, particularly the use of machine learning for forecasting academic outcomes. The trend continues in 2021, where terms like educational data mining, classification, and dropout become more prominent. This points to the field's increasing concern with early-warning systems and predictive interventions, especially for at-risk students. The rise in methodological terms like support vector and neural network also reflects greater algorithmic sophistication. In 2022, the cloud becomes more diversified, incorporating terms such as covid, early prediction, and academic performance. The appearance of pandemic-related terminology indicates the field's responsiveness to global disruptions, with a shift toward remote learning analytics and early performance forecasting. 2023 demonstrates a balanced integration of application-oriented and technical terms. The persistent presence of student, academic performance, and data mining is complemented by a methodological deepening with terms like regression and model. This suggests a maturing research agenda focused on refining predictive accuracy and model interpretability. By 2024, the vocabulary stabilizes, with consistent recurrence of core themes—machine learning, academic performance, student, prediction—and the introduction of more abstract concepts like artificial intelligence. The prominence of these terms implies a consolidation of the field, with attention shifting from proof-of-concept studies toward scalable, integrated systems in educational ecosystems.

A collage of words

AI-generated content may be incorrect.

Figure 2: Word cloud analysis through the years.

As we can observe in Figure 3, the distribution of cited reference counts among the articles presents a notable variation, illustrating a substantial spread within the interquartile range. The majority of articles receive a moderate number of citations, but several outliers represent publications with significantly higher citation counts. This indicates that a few studies attract disproportionate scholarly attention. The observations corroborate the methodology used in WoS Citation Reports (*Web of Science Core Collection: Finding the Average Number of Citations per Article in a Journal*, 2022), where the average citations per item provide an essential measure of scholarly impact.

A diagram of a box plot

AI-generated content may be incorrect.

Figure 3: Cited reference counts. Averaged citated reference count was 44.45, with standard deviation 25.87.

As illustrated in Figure 4, the number of publications in this field has been steadily increasing over the years, reflecting growing interest within the academic community. However, it is important to note that not all publications explicitly use "machine learning" as a keyword, as depicted in Figure 5. This discrepancy suggests that a significant portion of the research may be related to ML concepts without directly referencing the term, which highlights the diverse terminology used in the field and the broad scope of research topics that intersect with ML methodologies.

A graph with blue and orange bars

AI-generated content may be incorrect.

Figure 4: Number of publications per year with highlighted publications using “machine learning” as a keyword.

The growing popularity of the subject is also reflected in the increasing total number of citations over the years, indicating a rising impact and recognition within the academic community. As more research is conducted and published in this field, the cumulative citation count continues to climb, as seen on Figure 5. Interestingly, while the total number of citations has seen significant growth, the average number of citations per year has remained relatively consistent, as pictured in Figure 6. This suggests that while the volume of research output is expanding, the influence and reach of individual publications are maintaining a steady rate of scholarly engagement and reference over time.

A graph of a bar chart

AI-generated content may be incorrect.

Figure 5: Total citations per year.

A graph of a bar chart

AI-generated content may be incorrect.

Figure 6: Average citations per year.

However, raw citation counts tend to favor older publications simply due to having more time to accumulate references. To address this, we normalized citation data by calculating the average number of citations per year since publication, enabling a more equitable comparison across years. The resulting trend, shown in Figure 7, reveals that more recent publications have a higher citation rate per year, suggesting not only an increase in publication volume but also a growing immediacy and relevance of research in this area.

A graph of a number of people

AI-generated content may be incorrect.

Figure 7: Normalized average citations per year.

The analysis of the geographic distribution of research in Figure 8 reveals a clear dominance of China, which produced the highest frequency of publication. Significant contributions were also observed from other countries, including Australia, the United States, and various European nations, although their output was substantially lower in comparison. An examination of international collaboration networks, however, shows a different pattern, as shown in Figure 9. The most extensive collaboration links are identified between a cluster of countries including Australia, India, and nations in Europe Russia, indicating these regions serve as central hubs in the collaborative network. No significant collaboration is noted by US or China.

A map of the world

AI-generated content may be incorrect.

Figure 8: Country productivity plot of ML publications.

A map of the world with different colored countries/regions

AI-generated content may be incorrect.

Figure 9: Collaboration network of ML publications.

## Discussion

The application of ML in predicting academic success has emerged as a transformative area of research, bridging the fields of education, data science, and psychology. This study highlights the increasing role of ML in predicting academic success, as reflected in the growing volume of research and citations in recent years. The bibliometric trend analysis reveals a sustained interest in this field, with a noticeable shift toward more sophisticated data-driven approaches to educational analysis. The findings underscore the potential of ML models to provide actionable insights for improving student outcomes, identifying at-risk students, and enhancing the overall quality of education.

One of the key observations is the diversification of research themes over time. Early studies predominantly focused on basic predictive analytics and classification techniques. However, recent work has expanded to incorporate more complex models and interdisciplinary approaches, integrating insights from psychology (Burman et al., 2021), sociology (Davies et al., 2021), and others. This trend suggests a maturing field that is increasingly leveraging the power of ML to address multifaceted challenges in education. This reflects a growing consensus on the importance of predictive models in understanding and enhancing academic performance. Additionally, the rising number of publications and citations indicates a broader acceptance and recognition of ML's value in educational research. With rapid technological advancements, AI will undoubtedly play a crucial role in educational sciences (Author 2, 2019). This is further reflected in the analysis of authorship and collaboration, which reveals a complex global network. Although China leads in terms of sheer publication volume, the most prominent collaboration ties are concentrated among Australia, India, and various European institutions. This suggests that while one country may dominate in research output, the global network of knowledge exchange is more distributed, with other regions forming critical collaborative hubs.

It is crucial to recognize that the terms "ML" and "AI" are increasingly overused and misused (J. J. Chen & Lin, 2024), often serving as buzzwords that lead to misconceptions about their capabilities and limitations. This loose application dilutes the precision of academic discourse, making it difficult to categorize and synthesize relevant research effectively. Many studies claiming to use ML or AI lack rigorous methodology, while others employing related concepts may not explicitly reference these terms due to varying terminologies. This inconsistency hampers research discoverability (Zhao et al., 2023) and risks overshadowing genuinely innovative work, highlighting the need for more standardized and precise terminology in academic publications.

Furthermore, the choice of databases significantly influences the scope and comprehensiveness of bibliometric analyses. Different databases, such as SCOPUS, WOS, Taylor & Francis, and specialized educational repositories, offer varying levels of coverage, indexing, and citation data (Mongeon & Paul-Hus, 2016), which can affect the identification of key studies and trends (Delgado‐Quirós et al., 2024). The discrepancies between databases highlight the importance of selecting appropriate sources to ensure a balanced and accurate representation of the research landscape. In our case, although the initial search included SCOPUS, Taylor & Francis, and university registers, we ultimately relied on WOS for the bibliometric analysis due to its standardized metadata, compatibility with bibliometric tools, and lower risk of duplication, which allowed for a more consistent and reproducible analysis.

## Conclusion

This paper’s tri-scope approach—a synthesis of 33 literature reviews, an analysis of the 10 most-cited WOS articles, and a 285-article bibliometric study—offers a comprehensive view of the application of machine learning in academic success prediction. The findings highlight the potential of ML to revolutionize educational practices, but also call for greater methodological rigor, standardized terminology, and interdisciplinary collaboration to fully realize its benefits.

## Limitations and future directions

While this study provides valuable insights into the application of ML for predicting academic success, several limitations must be acknowledged. Firstly, the reliance on bibliometric data from only the selected database may introduce selection bias, as they might not capture the entirety of relevant publications, especially those published in regional or non-indexed journals. Similarly, the focus on specific keywords in the search strategy could have excluded studies employing alternative terminologies, thereby limiting the comprehensiveness of the analysis.

Another limitation pertains to the variability in the quality and consistency of metadata across different publications. Inconsistencies in keyword usage, authorship details, and citation practices can affect the reliability of bibliometric indicators. Furthermore, the dynamic nature of the field means that recent publications, may not yet have accumulated sufficient citations to accurately reflect their impact, leading to potential underestimation of their significance.

Future research should address these gaps by incorporating broader databases and more inclusive search strategies. Advanced text-mining techniques could uncover latent trends and themes, while longitudinal studies could track the evolution of ML applications in education over time.

Integrating bibliometric analysis provide deeper insights into the relevance of key studies. Collaboration between researchers, educators, and policymakers is essential to ensure ML models are effectively applied to enhance educational outcomes and support student success.

## Study registration and data availability

The study was registered in the OSF repository, where also data is available <https://osf.io/24wmr/?view_only=5dffcb1a58d946a2be437b88dcbbc2d7>

## References

Abdelkader, H. E., Gad, A. G., Abohany, A. A., & Sorour, S. E. (2022). An Efficient Data Mining Technique for Assessing Satisfaction Level With Online Learning for Higher Education Students During the COVID-19. *IEEE Access*, *10*, 6286–6303. https://doi.org/10.1109/ACCESS.2022.3143035

Abdul Bujang, S. D., Selamat, A., Krejcar, O., Mohamed, F., Cheng, L. K., Chiu, P. C., & Fujita, H. (2023). Imbalanced Classification Methods for Student Grade Prediction: A Systematic Literature Review. *IEEE Access*, *11*, 1970–1989. https://doi.org/10.1109/ACCESS.2022.3225404

Acosta-Gonzaga, E. (2023). The Effects of Self-Esteem and Academic Engagement on University Students’ Performance. *Behavioral Sciences*, *13*(4), 348. https://doi.org/10.3390/bs13040348

Alalawi, K., Athauda, R., & Chiong, R. (2023). Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review. *Engineering Reports*, *5*(12), e12699. https://doi.org/10.1002/eng2.12699

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, *37*, 13–49. https://doi.org/10.1016/j.tele.2019.01.007

Alghamdi, S., Soh, B., & Li, A. (2025). A Comprehensive Review of Dropout Prediction Methods Based on Multivariate Analysed Features of MOOC Platforms. *Multimodal Technologies and Interaction*, *9*(1), 3. https://doi.org/10.3390/mti9010003

Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., & Ali, N. (2022). Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance. *Computational Intelligence and Neuroscience*, *2022*, 1–11. https://doi.org/10.1155/2022/4151487

Alshanqiti, A., & Namoun, A. (2020). Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification. *IEEE Access*, *8*, 203827–203844. https://doi.org/10.1109/ACCESS.2020.3036572

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(1), 3. https://doi.org/10.1186/s41239-020-0177-7

Balaji, P., Alelyani, S., Qahmash, A., & Mohana, M. (2021). Contributions of Machine Learning Models towards Student Academic Performance Prediction: A Systematic Review. *Applied Sciences*, *11*(21), Article 21. https://doi.org/10.3390/app112110007

Bertolini, R., Finch, S. J., & Nehm, R. H. (2021). Testing the Impact of Novel Assessment Sources and Machine Learning Methods on Predictive Outcome Modeling in Undergraduate Biology. *Journal of Science Education and Technology*, *30*(2), 193–209. https://doi.org/10.1007/s10956-020-09888-8

Bregant, B. (2025). *Statistics code of the article* [Repository]. Github. https://github.com/borbregant/Doktorat\_all/blob/main/raziskovalni\_seminar/obdelava.ipynb

Bregant, B., Doz, D., & Hudovernik, S. (2025). Factors influencing tandem learning in mathematics. International Journal of Instruction, 18(1), 437–463.

Buitrago-Ropero, M. E., Ramírez-Montoya, M. S., & Laverde, A. C. (2023). Digital footprints (2005–2019): A systematic mapping of studies in education. *Interactive Learning Environments*, *31*(2), 876–889. https://doi.org/10.1080/10494820.2020.1814821

Burman, I., Som, S., & Hossain, S. A. (2021). Model for Analyzing Psychological Parameters Recommending Student Learning Behavior using Machine Learning. *International Journal of Computing and Digital Systems*, *10*(1), 973–990. https://doi.org/10.12785/ijcds/100188

Cankaya, B., Roberts, R., Douglas, S., Vigness, R., & Oztekin, A. (2024). What postpones degree completion? Discovering key predictors of undergraduate degree completion through explainable artificial intelligence (XAI). *Journal of Marketing Analytics*. https://doi.org/10.1057/s41270-024-00290-6

Chen, J., Fang, B., Zhang, H., & Xue, X. (2022). A systematic review for MOOC dropout prediction from the perspective of machine learning. *Interactive Learning Environments*, 1–14. https://doi.org/10.1080/10494820.2022.2124425

Chen, J. J., & Lin, J. C. (2024). Artificial intelligence as a double-edged sword: Wielding the POWER principles to maximize its positive effects and minimize its negative effects. *Contemporary Issues in Early Childhood*, *25*(1), 146–153. https://doi.org/10.1177/14639491231169813

Davies, H. C., Eynon, R., & Salveson, C. (2021). The Mobilisation of AI in Education: A Bourdieusean Field Analysis. *Sociology*, *55*(3), 539–560. https://doi.org/10.1177/0038038520967888

Delgado‐Quirós, L., Aguillo, I. F., Martín‐Martín, A., López‐Cózar, E. D., Orduña‐Malea, E., & Ortega, J. L. (2024). Why are these publications missing? Uncovering the reasons behind the exclusion of documents in free‐access scholarly databases. *Journal of the Association for Information Science and Technology*, *75*(1), 43–58. https://doi.org/10.1002/asi.24839

Dol, S. M., & Jawandhiya, P. M. (2024). Systematic Review and Analysis of EDM for Predicting the Academic Performance of Students. *Journal of The Institution of Engineers (India): Series B*, *105*(4), 1021–1071. https://doi.org/10.1007/s40031-024-00998-0

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic Review on Educational Data Mining. *IEEE Access*, *5*, 15991–16005. https://doi.org/10.1109/ACCESS.2017.2654247

Edgerton, E., & McKechnie, J. (2023). The relationship between student’s perceptions of their school environment and academic achievement. *Frontiers in Psychology*, *13*, 959259. https://doi.org/10.3389/fpsyg.2022.959259

Fahd, K., Venkatraman, S., Miah, S. J., & Ahmed, K. (2022). Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature. *Education and Information Technologies*, *27*(3), 3743–3775. https://doi.org/10.1007/s10639-021-10741-7

Gouseti, A., James, F., Fallin, L., & Burden, K. (2024). The ethics of using AI in K-12 education: A systematic literature review. *Technology, Pedagogy and Education*, 1–22. https://doi.org/10.1080/1475939X.2024.2428601

Guanin-Fajardo, J. H., Guaña-Moya, J., & Casillas, J. (2024). Predicting Academic Success of College Students Using Machine Learning Techniques. *Data*, *9*(4), 60. https://doi.org/10.3390/data9040060

Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., Wild, J., Krauss, S., Goretzko, D., & Stachl, C. (2021). Machine learning for the educational sciences. *Review of Education*, *9*(3), Article 3. https://doi.org/10.1002/rev3.3310

Istenič Starčič, A. (2019). Human learning and learning analytics in the age of artificial intelligence. *British Journal of Educational Technology*, *50*(6), 2974–2976. https://doi.org/10.1111/bjet.12879

Jang, Y., Choi, S., Jung, H., & Kim, H. (2022). Practical early prediction of students’ performance using machine learning and eXplainable AI. *Education and Information Technologies*, *27*(9), 12855–12889. https://doi.org/10.1007/s10639-022-11120-6

Jin, X. (2023). Predicting academic success: Machine learning analysis of student, parental, and school efforts. *Asia Pacific Education Review*. https://doi.org/10.1007/s12564-023-09915-4

Kassaw, C., & Demareva, V. (2023). Determinants of academic achievement among higher education student found in low resource setting, A systematic review. *PLOS ONE*, *18*(11), e0294585. https://doi.org/10.1371/journal.pone.0294585

Kim, K., & Kwon, K. (2024). A systematic review of the evaluation in K-12 artificial intelligence education from 2013 to 2022. *Interactive Learning Environments*, 1–29. https://doi.org/10.1080/10494820.2024.2335499

Liang, G., Jiang, C., Ping, Q., & Jiang, X. (2023). Academic performance prediction associated with synchronous online interactive learning behaviors based on the machine learning approach. *Interactive Learning Environments*, 1–16. https://doi.org/10.1080/10494820.2023.2167836

Lim, W. M., & Kumar, S. (2024). Guidelines for interpreting the results of bibliometric analysis: A sensemaking approach. *Global Business and Organizational Excellence*, *43*(2), 17–26. https://doi.org/10.1002/joe.22229

Lynam, S., Cachia, M., & Stock, R. (2024). An evaluation of the factors that influence academic success as defined by engaged students. *Educational Review*, *76*(3), 586–604. https://doi.org/10.1080/00131911.2022.2052808

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Namoun, A., & Alshanqiti, A. (2020). Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. *Applied Sciences*, *11*(1), 237. https://doi.org/10.3390/app11010237

Ncube, M. M., & Ngulube, P. (2024). Optimising Data Analytics to Enhance Postgraduate Student Academic Achievement: A Systematic Review. *Education Sciences*, *14*(11), 1263. https://doi.org/10.3390/educsci14111263

Nti, I. K., Umar Bawah, F., Quarcoo, J. A., & Kalos, F. (2022). A Bibliometric Analysis of Soft Computing Technology Applications Trends and Characterisation in Educational Research: Africa. *Africa Education Review*, *19*(3), 55–77. https://doi.org/10.1080/18146627.2023.2284744

Page M. J., McKenzie J. E., Bossuyt P. M., Boutron I., Hoffmann T. C., Mulrow C. D., Moher D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, 372. https://doi.org/10.1136/bmj.n71

Pektaş, Ş. T. (2023). A Systematic Analysis of Machine Learning Studies in Education. *The 15th International Conference on Education Technology and Computers*, 451–455. https://doi.org/10.1145/3629296.3629368

Pelima, L. R., Sukmana, Y., & Rosmansyah, Y. (2024). Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review. *IEEE Access*, *12*, 23451–23465. https://doi.org/10.1109/ACCESS.2024.3361479

Pinto, A. S., Abreu, A., Costa, E., & Paiva, J. (2023). How Machine Learning (ML) is Transforming Higher Education: A Systematic Literature Review. *Journal of Information Systems Engineering and Management*, *8*(2), 21168. https://doi.org/10.55267/iadt.07.13227

Pires, J. P. J., Brito Correia, F., Gomes, A., Borges, A. R., & Bernardino, J. (2024). Predicting Student Performance in Introductory Programming Courses. *Computers*, *13*(9), 219. https://doi.org/10.3390/computers13090219

Pradana, M., Elisa, H. P., & Syarifuddin, S. (2023). Discussing ChatGPT in education: A literature review and bibliometric analysis. *Cogent Education*, *10*(2), 2243134. https://doi.org/10.1080/2331186X.2023.2243134

PRISMA-P Group, Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1. https://doi.org/10.1186/2046-4053-4-1

Rahul, & Katarya, R. (2023). A Systematic Review on Predicting the Performance of Students in Higher Education in Offline Mode Using Machine Learning Techniques. *Wireless Personal Communications*, *133*(3), 1643–1674. https://doi.org/10.1007/s11277-023-10838-x

Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and Predicting Students’ Performance by Means of Machine Learning: A Review. *Applied Sciences*, *10*(3), Article 3. https://doi.org/10.3390/app10031042

Schillereff, D., Clarke, L., Shuttleworth, E., & Alderson, D. (2023). Evaluating success in a changing academic landscape. *Earth Surface Processes and Landforms*, *48*(12), 2387–2394. https://doi.org/10.1002/esp.5634

Sekeroglu, B., Abiyev, R., Ilhan, A., Arslan, M., & Idoko, J. B. (2021). Systematic Literature Review on Machine Learning and Student Performance Prediction: Critical Gaps and Possible Remedies. *Applied Sciences*, *11*(22), 10907. https://doi.org/10.3390/app112210907

Shafiq, D. A., Marjani, M., Habeeb, R. A. A., & Asirvatham, D. (2022). Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review. *IEEE Access*, *10*, 72480–72503. https://doi.org/10.1109/ACCESS.2022.3188767

Tan, B., Jin, H.-Y., & Cutumisu, M. (2024). The applications of machine learning in computational thinking assessments: A scoping review. *Computer Science Education*, *34*(2), 193–221. https://doi.org/10.1080/08993408.2023.2245687

Tete, M. F., Sousa, M. D. M., De Santana, T. S., & Silva, S. F. (2022). Aplicação de métodos preditivos em evasão no ensino superior: Uma revisão sistemática da literatura. *Education Policy Analysis Archives*, *30*. https://doi.org/10.14507/epaa.30.6845

Umer, R., Susnjak, T., Mathrani, A., & Suriadi, L. (2023). Current stance on predictive analytics in higher education: Opportunities, challenges and future directions. *Interactive Learning Environments*, *31*(6), 3503–3528. https://doi.org/10.1080/10494820.2021.1933542

Urdaneta-Ponte, M. C., Mendez-Zorrilla, A., & Oleagordia-Ruiz, I. (2021). Recommendation Systems for Education: Systematic Review. *Electronics*, *10*(14), 1611. https://doi.org/10.3390/electronics10141611

*Web of Science Core Collection: Finding the average number of citations per article in a journal*. (2022). Clarivate. https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-Core-Collection-Finding-the-average-number-of-citations-per-article-in-a-journal?language=en\_US

Xiao, W., & Hu, J. (2023). A state‐of‐the‐art survey of predicting students’ performance using artificial neural networks. *Engineering Reports*, *5*(8), e12652. https://doi.org/10.1002/eng2.12652

York, T. T., Gibson, C., & Rankin, S. (2015). Defining and Measuring Academic Success. *Practical Assessment, Research, and Evaluation*, *20*(1), 5. https://doi.org/10.7275/HZ5X-TX03

Zhao, L., Chen, K., Song, J., Zhu, X., Sun, J., Caulfield, B., & Namee, B. M. (2021). Academic Performance Prediction Based on Multisource, Multifeature Behavioral Data. *IEEE Access*, *9*, 5453–5465. https://doi.org/10.1109/ACCESS.2020.3002791

Zhao, L., Shen, C., Liu, M., Zhang, J., Cheng, L., Li, Y., Yuan, L., Zhang, J., & Tian, J. (2023). Comparison of Reporting and Transparency in Published Protocols and Publications in Umbrella Reviews: Scoping Review. *Journal of Medical Internet Research*, *25*, e43299. https://doi.org/10.2196/43299

## Appendix

### Most cited studies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Author Full Names** | **Article Title** | **Source Title** | **Cited Reference Count** |
| 1 | Jang, Yeonju; Choi, Seongyune; Jung, Heeseok; Kim, Hyeoncheol | Practical early prediction of students' performance using machine learning and eXplainable AI | EDUCATION AND INFORMATION TECHNOLOGIES | 143 |
| 2 | Fahd, Kiran; Venkatraman, Sitalakshmi; Miah, Shah J.; Ahmed, Khandakar | Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature | EDUCATION AND INFORMATION TECHNOLOGIES | 117 |
| 3 | Bertolini, Roberto; Finch, Stephen J.; Nehm, Ross H. | Testing the Impact of Novel Assessment Sources and Machine Learning Methods on Predictive Outcome Modeling in Undergraduate Biology | JOURNAL OF SCIENCE EDUCATION AND TECHNOLOGY | 116 |
| 4 | Puyana-Romero, Virginia; Larrea-Alvarez, Cesar Marcelo; Diaz-Marquez, Angela Maria; Hernandez-Molina, Ricardo; Ciaburro, Giuseppe | Developing a Model to Predict Self-Reported Student Performance during Online Education Based on the Acoustic Environment | SUSTAINABILITY | 105 |
| 5 | Abdelkader, Hanan E.; Gad, Ahmed G.; Abohany, Amr A.; Sorour, Shaymaa E. | An Efficient Data Mining Technique for Assessing Satisfaction Level With Online Learning for Higher Education Students During the COVID-19 | IEEE ACCESS | 97 |
| 6 | Gaftandzhieva, Silvia; Talukder, Ashis; Gohain, Nisha; Hussain, Sadiq; Theodorou, Paraskevi; Salal, Yass Khudheir; Doneva, Rositsa | Exploring Online Activities to Predict the Final Grade of Student | MATHEMATICS | 96 |
| 7 | Rico-Juan, Juan Ramon; Cachero, Cristina; Macia, Hermenegilda | Study regarding the influence of a student's personality and an LMS usage profile on learning performance using machine learning techniques | APPLIED INTELLIGENCE | 95 |
| 8 | Guanin-Fajardo, Jorge Humberto; Guana-Moya, Javier; Casillas, Jorge | Predicting Academic Success of College Students Using Machine Learning Techniques | DATA | 93 |
| 9 | Estrada Molina, Odiel; Fuentes Cancel, Dieter Reynaldo | Is it possible to predict academic performance? An analysis from educational technology | REVISTA FUENTES | 89 |
| 10 | Waheed, Hajra; Hassan, Saeed-Ul; Nawaz, Raheel; Aljohani, Naif R.; Chen, Guanliang; Gasevic, Dragan | Early prediction of learners at risk in self-paced education: A neural network approach | EXPERT SYSTEMS WITH APPLICATIONS | 88 |

Table : Most cited studies.

### Existing literature review studies

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No.** | **Authors,  Year of publication, Title,  Journal** | **Review methodology, No. of studies included and period (if not given, there were no strict timelines)** | **Level of prediction: degree level (obtaining a degree/graduation), year level, course level, exam level** | **ML or other prediction techniques** | **RQ, Findings & Gaps identified** |
| 1 | Pelima et al., 2024; Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review;  IEEEAccess | SLR methodology based on Kitchenham’s guidelines; 70 journal articles (2018-2023). | Degree (graduation/dropout), year (at-risk students), course (performance), exam (test results). | SVM, RF, ANN/MLP, LR, k-NN. | RQ1: State-of-the-art EDM/EDA research. RQ2: Data sources & techniques. RQ3: Gaps & challenges.  High ML accuracy (~90%). LMS & SIS are key data sources. Academic, behavioral, and demographic data matter.  Data limitations, model interpretability, ethical concerns, limited real-world application. |
| 2 | Umer et al., 2023; Current stance on predictive analytics in higher  education: opportunities, challenges and future  directions; Interactive Learning Environments | SLR based on PRISMA guidelines; 40 journal articles (2008-2018) | Course (pass/fail, dropout risk), degree (graduation likelihood) | DT, RF, SVM, LGR, NB, KNN, Ensemble methods. | RQ1: Types of data used. RQ2: ML methods for student performance prediction. RQ3: Evaluation measures. RQ4: Challenges & limitations. RQ5: Future research directions.  LMS & SMS data are key sources. Pre-academic, demographic, and assessment data are strong predictors. ML models achieve high accuracy.  Lack of benchmark datasets, generalizability issues, small sample sizes, limited real-world implementation, ethical concerns in data use. |
| 3 | Shafiq et al., 2022; Student Retention Using Educational Data Mining  and Predictive Analytics: A Systematic  Literature Review; IEEEAccess | SLR based on PRISMA guidelines; 100 articles (2017-2021) | Student retention, dropout risk, academic success. | Supervised ML (RF, DT, LR, NB), DL (ANN, MLP), Ensemble methods, Unsupervised Learning (Clustering). | RQ1: Factors influencing student retention. RQ2: Learning Analytics approaches. RQ3: ML algorithms for retention prediction;  Traditional academic factors (grades, attendance) dominate prediction. Emerging factors: behavioral and educator-related attributes. Supervised ML is widely used;  Lack of benchmark datasets, limited generalizability, ethical concerns, imbalance in datasets, underuse of clustering methods. |
| 4 | Alalawi et al., 2023; Contextualizing the current state of research on the use of  machine learning for student performance prediction: A  systematic literature review; Engineering Reports | SLR using Kitchenham’s approach; 162 articles (2010-2022). | Degree (graduation/dropout), course (performance), assessment (exam results). | Classification (138 studies), Regression (25 studies), Clustering (9 studies). Most used: DT, RF, NB, ANN, SVM | RQ1: Goals of ML in predicting student performance. RQ2: Common ML methods. RQ3: Key features for prediction. RQ4: Feature selection techniques. RQ5: Learning environments where ML is applied. RQ6: ML tools and platforms used. RQ7: Actions taken for at-risk students;  Historical academic data (grades, LMS interactions) is the most used predictor. Supervised learning (classification) is the dominant approach. WEKA, Python, and R are the most used platforms;  Lack of studies on real-world interventions for at-risk students. Limited use of clustering techniques. Small datasets and generalizability issues. Ethical concerns in data usage. |
| 5 | Alyahyan & Düştegör, 2020; Predicting academic success in higher  education: literature review and best  practices; International Journal of Educational Technology in Higher  Education | SLR of key studies on data mining techniques for predicting academic success; 17 studies (2015-2020). | Degree (graduation likelihood), year (academic success per year), course (performance in a specific course). | Classification (DT, NB, SVM, ANN), Regression, Clustering. | RQ1: How is academic success defined and measured? RQ2: What are the key factors influencing academic success? RQ3: What ML techniques are used for prediction? RQ4: What are the best practices for applying data mining to education?;  Prior academic achievement and student demographics are the most influential factors. Supervised learning (classification) is the most used approach. Including university data improves prediction accuracy. Limited studies on interventions for at-risk students, lack of benchmark datasets, generalizability issues, underuse of clustering techniques. |
| 6 | Balaji et al., 2021; Contributions of Machine Learning Models towards Student  Academic Performance Prediction: A Systematic Review; Applied Sciences | SLR; 56 articles (filtered from 2700 initially considered). | Degree (graduation/dropout), course (performance), year (academic success per year). | DT, RF, NB, SVM, ANN, ensemble methods | RQ1: ML models used for predicting student performance. RQ2: Estimation methods & performance metrics. RQ3: Datasets and collection methods. RQ4: Features used for prediction. RQ5: Model comparisons for reliability;  ML models can accurately predict student performance. Ensemble models (32%) and Decision Trees (26%) are widely used. Accuracy, precision, recall, and F1-score are common evaluation metrics. Datasets include academic, demographic, and behavioral features;  Lack of benchmark datasets, limited real-world applications, feature selection inconsistencies, and generalizability issues. |
| 7 | Shafiq et al., 2022; Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review, IEEEAccess | SLR on student retention using EDM and Predictive Analytics; 78 studies (2015–2022). | Degree (graduation/dropout), year (academic progression and retention). | DT, RF, SVM, LGR, NB, ANN, Ensemble methods | RQ1: Factors influencing student retention. RQ2: ML models used for retention prediction. RQ3: Evaluation metrics and datasets used. RQ4: Challenges in retention prediction. RQ5: Future research directions;  Academic performance (GPA, attendance), behavioral data, and demographic factors influence retention. Supervised ML models (DT, RF, ANN) are dominant. LMS and SIS data are key sources,  Limited real-world implementation, small sample sizes, lack of standardized datasets, ethical concerns in predictive analytics. |
| 8 | Fahd et al., 2022; Application of machine learning in higher education  to assess student academic performance, at-risk,  and attrition: A meta-analysis of literature, Education and Information Technologies | SLR and meta-analysis using the PRISMA framework; 89 studies (2010–2020). | Degree (graduation/dropout), year (academic success per year), at-risk students, attrition prediction. | Supervised Learning (DT, RF, SVM, NB, ANN, LGR), Unsupervised Learning (Clustering, K-Means), Reinforcement Learning (Q-Learning, SARSA). | RQ1: ML models used for predicting student performance and retention. RQ2: Evaluation metrics applied. RQ3: Common datasets and features. RQ4: ML applications in at-risk student prediction. RQ5: Future research directions; Limited real-world implementation, lack of benchmark datasets, underutilization of clustering and reinforcement learning, ethical concerns in predictive analytics. |
| 9 | Namoun & Alshanqiti, 2020; Predicting Student Performance Using Data Mining and  Learning Analytics Techniques: A Systematic Literature Review, Applied Sciences | SLR using PICO and PRISMA frameworks; 62 studies (2010–2020). | Course (performance based on learning outcomes), program (graduation likelihood). | Regression models, Supervised Learning (DT, ANN, Naïve Bayes, SVM), Unsupervised Learning (Clustering), Hybrid Models. | RQ1: How is academic performance measured using learning outcomes? RQ2: What intelligent models predict student performance? RQ3: What are the key predictors of academic performance?  Student learning outcomes are strong indicators of academic success. Online learning activities, assessment data, and academic emotions are key predictors. Supervised ML models (Regression, DT, ANN) dominate.  Limited research on interventions for at-risk students, lack of standardized datasets, minimal use of clustering methods, challenges in real-world implementation. |
| 10 | Pektaş, 2023; A Systematic Analysis of Machine Learning Studies in Education, ICETC 2023: The 15th International Conference on Education Technology and Computers | Bibliometric analysis using keyword co-occurrence and network visualization. Data sourced from Web of Science; 628 studies (1979–2023) | Profiling and prediction (dropout risk, academic performance, admissions). Assessment (automated grading, feedback). Distance learning success. | Supervised Learning (DT, SVM, ANN), Unsupervised Learning (Clustering), NLP for text analysis. | RQ1: Trends in ML publications in education. RQ2: Key research themes in ML for education. RQ3: Gaps in current ML-based educational research;  ML is mainly used for profiling, assessment, intelligent tutoring systems, MOOCs, NLP applications, and distance learning prediction. Growth in publications since 2017;  Overemphasis on applications rather than ethical, pedagogical, and socio-cultural aspects. Limited theoretical frameworks guiding ML studies. |
| 11 | Rahul & Katarya, 2023; A Systematic Review on Predicting the Performance  of Students in Higher Education in Ofine Mode Using  Machine Learning Techniques, Wireless Personal Communications | SLR on predicting student performance in offline education using ML techniques; 109 studies (2010–2020). | Degree (graduation likelihood), course (performance), year (academic success per year). | Supervised Learning (DT, SVM, ANN, NB, LGR), Unsupervised Learning (Clustering, K-Means), Hybrid Models. | RQ1: Most used ML techniques for student performance prediction. RQ2: Datasets and features used. RQ3: Key evaluation metrics. RQ4: Trends in ML applications for student success;  Student performance is influenced by demographic, academic, and behavioral factors. Classification-based models (DT, RF, ANN) dominate. Accuracy, precision, and F1-score are commonly used metrics;  Lack of standardized datasets, generalizability issues, limited real-world implementation, underuse of clustering methods. |
| 12 | Sekeroglu et al., 2021; Systematic Literature Review on Machine Learning and Student  Performance Prediction: Critical Gaps and Possible Remedies, Applied Sciences | SLR using PRISMA framework; 176 studies (2010–2020). | Degree (graduation/dropout), course (performance), year (academic progression), at-risk students. | Supervised Learning (DT, RF, SVM, ANN, LGR), Unsupervised Learning (Clustering), Deep Learning (DNN, RNN), Hybrid Models. | RQ1: ML models used in student performance prediction. RQ2: Evaluation metrics and validation strategies. RQ3: Datasets and features used. RQ4: Standardization challenges in ML applications for education;  Supervised models dominate; ANN, SVM, and RF widely used. Online learning data improves prediction. Accuracy, F1-score, and ROC AUC are key evaluation metrics. Cross-validation is preferred for validation;  Lack of standardized datasets, generalizability issues, limited real-world applications, underuse of clustering methods, and ethical concerns in predictive analytics. |
| 13 | J. Chen et al., 2022; A systematic review for MOOC dropout prediction from  the perspective of machine learning Interactive Learning Environments | SLR using the PRISMA protocol; 78 studies (2012–2022). | MOOC dropout prediction at the course level. | Supervised Learning (DT, SVM, ANN, NB, LGR), Deep Learning (CNN, LSTM, Transformers), Hybrid Models. | RQ1: Factors influencing MOOC dropout. RQ2: ML methods for dropout prediction. RQ3: Feature extraction techniques. RQ4: Key challenges in dropout prediction;  Clickstream and text-based features improve dropout prediction. Deep learning models show promising results but require large datasets. Accuracy, AUC, and F1-score are key evaluation metrics;  Interpretability issues in deep learning models, imbalanced datasets, lack of standardized dropout definitions, and challenges in modeling semantic learning trajectories. |
| 14 | Tan et al., 2024; The applications of machine learning in computational thinking assessments: a scoping review; Computer Science Education | Scoping Review using PRISMA and Arksey & O’Malley’s framework; 20 studies | Assessment of Computational Thinking (CT) at different levels (coursework, projects, and student interactions). | Regression (Linear, Lasso, Ridge), Classification (NB, SVM, DT, FR, ANN), Unsupervised Learning (Clustering, NLP), No reinforcement learning identified. | RQ1: Educational contexts of ML-based CT assessments. RQ2: Characteristics of datasets used. RQ3: ML models applied in CT assessments. RQ4: Aspects of CT measured;  ML is increasingly applied to automate CT assessments. CT competencies are mostly assessed in programming-related environments. Small datasets are a common limitation. Evaluation metrics include accuracy, precision, recall, and F1-score;  Lack of standardized assessment frameworks, small sample sizes, underuse of reinforcement learning, and limited application in non-STEM subjects. |
| 15 | Kim & Kwon, 2024; A systematic review of the evaluation in K-12 artificial intelligence education from 2013 to 2022, Interactive Learning Environments | SLR using Kitchenham’s framework; 36 studies (2013–2022). | Evaluations in K-12 AI education (cognitive & affective learning outcomes). | No direct ML prediction; focuses on AI education evaluation methods (knowledge tests, surveys, qualitative assessments). | RQ1: Evaluation contexts, types, and research designs in K-12 AI education. RQ2: Types of learning outcomes evaluated. RQ3: Evaluation methods used to measure AI learning;  Evaluations focus on AI literacy, ML concepts, and student perceptions. Most studies use surveys and knowledge tests. Informal learning environments (workshops, summer camps) dominate over formal classroom settings;  Overreliance on self-report surveys and summative assessments. Limited evaluation of emotional and behavioral aspects. Need for more formative assessments and studies in formal K-12 settings. |
| 16 | Pradana et al., 2023; Discussing ChatGPT in Education: A Literature Review and Bibliometric Analysis, Cogent Education | SLR and Bibliometric Analysis; 93 articles (2022-2023) | Not focused on student performance prediction but rather on ChatGPT’s role in education. Article features citation prediction | Gradient Boosting DT, SVM, XGBoost | RQ1: Research trends on ChatGPT in education. RQ2: Key contributors and topics. RQ3: Research gaps and future directions;  Growing interest in ChatGPT in education since 2022. Key themes include AI's role in teaching and learning. Gaps exist in studying challenges, teaching applications, and knowledge implications. |
| 17 | Buitrago-Ropero et al., 2023; Digital Footprints (2005–2019): A Systematic Mapping of Studies in Education; Interactive Learning Environments | Systematic Mapping Study; 46 articles (2005-2019) | School success/failure, learning analysis, dropout prediction, psychometric modeling | ML, Data Analytics, Big Data, Educational Data Mining (EDM), Learning Analytics | RQ1: Conceptions of digital footprints. RQ2: Educational activities and processes where digital footprints are used. RQ3: Resources for collecting digital footprints. RQ4: Technologies used for analyzing digital footprints;  Digital footprints are widely used for learning analytics, psychometric modeling, and predicting school success. There is limited research on digital footprints in MOOCs and ethical concerns about data privacy. |
| 18 | Nti et al., 2022; A Bibliometric Analysis of Soft Computing Technology Applications Trends and Characterisation in Educational Research: Africa, Africa Education Review | Bibliometric analysis using PRISMA model; 1,358 articles (1960-2021) | Student academic performance prediction, learning analytics, technology-enhanced education | ML in general, EDM, AI, Soft Computing | RQ1: Trends in SCTAE research in Africa. RQ2: Most impactful authors, papers, and institutions. RQ3: Intellectual framework of SCTAE knowledge base. RQ4: Key research areas and challenges in SCTAE;  Growth in AI and machine learning for education. Focus on student academic performance and learning analytics. Limited research on early education and AI-driven student assessments. Challenges in network infrastructure, educator training, and ethical AI use. |
| 19 | Gouseti et al., 2024; The Ethics of Using AI in K-12 Education: A Systematic Literature Review, Technology, Pedagogy and Education | SLR based on PRISMA methodology; 25 peer-reviewed studies (2010-2023) | Not focused on student performance prediction but on ethical concerns in AI use in K-12 education | AI-based adaptive learning, ITS, Emotion AI, AI-powered surveillance, ML in general | RQ1: Challenges relating to AI ethics in K-12 education. RQ2: Key responses to AI ethics in schools;  AI ethics is under-researched in K-12 education. Teachers and students lack awareness of AI ethics. Ethical concerns include bias, surveillance, privacy, emotional AI, and algorithmic decision-making. Need for AI literacy, ethics integration in curricula, and professional development for teachers. |
| 20 | Alghamdi et al., 2025; A Comprehensive Review of Dropout Prediction Methods Based on Multivariate Analysed Features of MOOC Platforms; Multimodal Technol. Interact. | SLR based on PRISMA methodology; 105 articles (2018-2024) | MOOC dropout prediction at the course level | Supervised Learning (DT, RF, SVM, ANN, NB, LGR), Deep Learning (CNN, LSTM, Transformers), Meta-heuristic approaches | RQ1: Factors influencing MOOC dropout. RQ2: ML and AI methods for dropout prediction. RQ3: Feature selection techniques. RQ4: Challenges and optimisation issues in dropout prediction. Clickstream and engagement features enhance dropout prediction. Deep learning models (CNN, LSTM) improve accuracy but require large datasets. Meta-heuristics aid feature selection. |
| 21 | Dutt et al., 2017; A Systematic Review on Educational Data Mining; IEEE Access | SLR based on Kitchenham’s methodology; 166 articles (1983–2016) | Degree (graduation likelihood), course (performance), student behavior clustering | Clustering (Hierarchical, K-Means, DBSCAN, Fuzzy C-Means), Classification (DT, NB, ANN), Association Rule Mining | RQ1: Applications of clustering in EDM. RQ2: How clustering differs from traditional DM methods. RQ3: Key predictors in EDM. RQ4: Future research directions. Clustering aids in understanding student performance, retention, and behavior. EDM requires specialized algorithms different from traditional DM methods. |
| 22 | Aldowah et al., 2019;Educational Data Mining and Learning Analytics for 21st Century Higher Education: A Review and Synthesis; Telematics and Informatics | SLR covering 402 studies (2000–2017) | Degree (graduation likelihood), course (performance), behavioral analytics | Classification (DT, NB, ANN), Clustering (K-Means, DBSCAN), Association Rule Mining, Visual Data Mining | RQ1: How EDM and LA techniques solve learning problems. RQ2: Best-suited DM techniques for educational challenges. EDM and LA improve personalized learning, dropout prediction, and curriculum design. Visualization aids decision-making. |
| 23 | Alhothali et al., 2022; Predicting Student Outcomes in Online Courses Using Machine Learning Techniques: A Review; Sustainability | SLR covering 67 studies (2017–2021) | Course (performance, dropout risk) | Supervised Learning (DT, RF, SVM, ANN, NB, LGR), Deep Learning (CNN, LSTM), Hybrid Models | RQ1: Strategies for learner outcome prediction. RQ2: Key predictive variables. RQ3: ML methodologies used. RQ4: Challenges in online learning prediction. Behavioral data (clickstream, engagement) are strong predictors. CNN and LSTM improve performance. Public datasets like OULAD and KDDcup are widely used. |
| 24 | Xiao & Hu, 2023; A State-of-the-Art Survey of Predicting Students' Performance Using Artificial Neural Networks; Engineering Reports | SLR covering 39 studies (2016–2021) | Course (performance), degree (graduation likelihood) | ANN (FNN, CNN, RNN, GNN), Deep Learning (LSTM, Attention Mechanisms) | RQ1: Objectives of ANN-based prediction. RQ2: Datasets and feature properties. RQ3: ANN types and structures. RQ4: Parameter and hyperparameter optimization. RQ5: Feature selection and representation learning. RQ6: Performance of ANN models. ANN achieves high accuracy (>90%) in student performance prediction. Representation learning (CNN, GNN) enhances predictive power. Most studies use private datasets. |
| 25 | Pinto et al., 2023; How Machine Learning (ML) is Transforming Higher Education: A Systematic Literature Review; Journal of Information Systems Engineering & Management | SLR using PRISMA methodology; 171 studies (2019–2023) | Degree (graduation likelihood), course (performance), employability prediction | Supervised Learning (RF, SVM, NB, ANN, LGR), Deep Learning (CNN, LSTM, MLP), Gradient Boosting (XGBoost, CatBoost, LGBM) | RQ1: Applications of ML in higher education. RQ2: Key ML techniques used. RQ3: Datasets and features in ML models. RQ4: Challenges in ML adoption in higher education. ML is widely applied in predicting student performance and employability. RF and SVM are most used in supervised learning. Python is the dominant programming language. |
| 26 | Abdul Bujang et al., 2023: Imbalanced Classification Methods for Student Grade Prediction: A Systematic Literature Review; IEEE Access | SLR based on Kitchenham’s guidelines; 43 studies (2015–2021) | Course (student grade prediction) | Supervised Learning (DT, RF, SVM, ANN, NB, LGR), Data-level balancing (SMOTE, ADASYN, RUS, Hybrid Methods), Algorithm-level (Cost-sensitive Learning), Hybrid Approaches | RQ1: Methods for handling imbalanced classification in student grade prediction. RQ2: Effectiveness of different balancing techniques. RQ3: Key datasets and evaluation metrics. SMOTE and ADASYN are widely used to balance datasets. Hybrid approaches improve accuracy in multi-class settings. |
| 27 | Alsariera et al., 2022; Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance; Computational Intelligence and Neuroscience | SLR covering 39 studies (2015–2021) | Course (academic performance) | Supervised Learning (DT, SVM, ANN, KNN, LinR, NB) | RQ1: Key predictive features in student performance. RQ2: Common ML algorithms used. RQ3: Performance and accuracy of ML models. ANN outperformed other models with high accuracy (~98%). Academic, demographic, internal assessment, and personal attributes are strong predictors. |
| 28 | Dol & Jawandhiya, 2024; Systematic Review and Analysis of EDM for Predicting the Academic Performance of Students; Journal of the Institution of Engineers (India) Series B | SLR covering 231 studies (2010–2020) | Degree (graduation likelihood), course (performance), student retention | Supervised Learning (DT, SVM, ANN, NB, LGR), Clustering (K-Means, DBSCAN), Association Rule Mining (Apriori), Text Mining, Sentiment Analysis, Process Mining, Graph Mining | RQ1: Common EDM techniques for academic performance prediction. RQ2: Classification and clustering techniques used. RQ3: Feature selection methods. RQ4: Challenges in EDM. Naïve Bayes, J48, and SVM are the most used classification algorithms. K-Means is the most used clustering technique. Data from LMS and university records are key predictors. |
| 29 | Tete et al., 2022; Predictive Models for Higher Education Dropout: A Systematic Literature Review; Education Policy Analysis Archives | SLR using PRISMA methodology; 48 studies (2010–2020) | Degree (dropout prediction) | Supervised Learning (DT, SVM, ANN, NB, LGR), Clustering (K-Means), Bayesian Networks, Logistic Regression | RQ1: Methodological and contextual characteristics of dropout prediction studies. RQ2: Determinant variables for dropout prediction. RQ3: Institutional actions to reduce dropout. Decision trees and Bayesian networks are the most used techniques. Socioeconomic, academic, psychological, and engagement factors strongly influence dropout risk. |
| 30 | Urdaneta-Ponte et al., 2021; Recommendation Systems for Education: Systematic Review; Electronics | SLR covering 98 studies (2015–2020) | Not focused on direct student performance prediction but on educational recommendation systems | Collaborative Filtering, Content-Based Filtering, Hybrid Models, Machine Learning (Neural Networks, Bayesian Techniques, Genetic Algorithms, Fuzzy Sets) | RQ1: Types of education and areas covered by RSs. RQ2: Target users of RSs. RQ3: Developmental approaches in RSs. RQ4: Platforms used in educational RSs. RSs are mainly used in formal education, focusing on recommending learning resources and courses. Hybrid models integrating ML techniques have increased in recent years. |
| 31 | Rastrollo-Guerrero et al., 2020; Analyzing and Predicting Students’ Performance by Means of Machine Learning: A Review; Applied Sciences | SLR covering 64 studies (2013–2019) | Degree (dropout risk), course (performance), student knowledge acquisition | Supervised Learning (DT, SVM, ANN, NB, LGR), Unsupervised Learning (K-Means, Hierarchical Clustering), Recommender Systems (Collaborative Filtering), Deep Learning (ELM, RNN, LSTM) | RQ1: Common ML techniques for predicting student performance. RQ2: Key predictive factors. RQ3: Impact of ML applications in education. SVM and RF are the most commonly used models for performance prediction. Recommender systems enhance learning engagement. Neural networks improve prediction accuracy but require large datasets. |
| 32 | Pires et al., 2024; Predicting Student Performance in Introductory Programming Courses; Computers | SLR based on Kitchenham’s methodology; 11 studies (2015–2024) | Course (performance in Introductory Programming) | Supervised Learning (DT, SVM, ANN, NB, LGR), Deep Learning (DNN, PNN), Hybrid Models | RQ1: Most-used ML algorithms for predicting performance in programming courses. RQ2: Datasets and evaluation metrics used. RQ3: ML algorithms with the best predictive accuracy. J48, Naïve Bayes, Random Forest, and SVM are the most frequently used models. Accuracy and F1-score are key evaluation metrics. Deep learning models show promise but are underutilized. |
| 33 | Ncube & Ngulube, 2024; Optimising Data Analytics to Enhance Postgraduate Student Academic Achievement: A Systematic Review; Education Sciences | SLR using PRISMA methodology; 26 studies (2016–2024) | Postgraduate (academic achievement, retention) | Descriptive, Predictive, Prescriptive Analytics, ML, Social Network Analysis, Text Analytics, Data Visualization | RQ1: Data analytics approaches used in postgraduate education. RQ2: Challenges in implementing data analytics. RQ3: Best practices for fostering a data-driven learning environment. Data analytics helps identify struggling students, personalize learning, and optimize teaching strategies. Predictive models aid in early intervention. |

Table : Analysis of the 33 review studies in the WOS, SCOPUS and Taylor and Francis collection. Legend of abbreviations: AI (Artificial Intelligence), ANN (Artificial Neural Networks), AUC (Area Under the Curve), CNN (Convolutional Neural Networks), CT (Computational Thinking), DL (Deep Learning), DT (Decision Trees), EDM (Educational Data Mining), EDA (Educational Data Analytics), F1-score (F1-score), GPA (Grade Point Average), ITS (Intelligent Tutoring Systems), k-NN (k-Nearest Neighbors), LGR (Logistic Regression), LMS (Learning Management System), LSTM (Long Short-Term Memory), ML (Machine Learning), MLP (Multilayer Perceptron), MOOC (Massive Open Online Course), NB (Naïve Bayes), NLP (Natural Language Processing), PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), Q-Learning (Q-Learning), RF (Random Forest), RNN (Recurrent Neural Networks), ROC (Receiver Operating Characteristic), SARSA (State-Action-Reward-State-Action), SCTAE (Soft Computing Technology Applications in Education), SLR (Systematic Literature Review), SMS (Student Management System), SVM (Support Vector Machine), WEKA (Waikato Environment for Knowledge Analysis), XGBoost (Extreme Gradient Boosting).