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

### Abstract

…

### 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). 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 as itself.

Student success is a key indicator of educational quality. York et al., (2015) defines academic success as a sum of six core components: academic achievement, satisfaction, skill acquisition, persistence, learning objectives, and career success, but is often assumed as grade point average only (Alyahyan & Düştegör, 2020). This corresponds to what students view as academic success i.e. a combination of outcomes including grades and more holistic outcomes of personal development and achievements (Lynam et al., 2024). It varies widely among individuals (Schillereff et al., 2023), and is influenced by a combinations of personal factors such as self-regulation, motivation, and self-esteem; environmental factors like socioeconomic status, school environment, and peer support; and lifestyle factors including dietary habits, sleep, and stress management (CITATI consensus what affects academic success https://consensus.app/results/?q=What%20affects%20academic%20success&sjr\_min=1&sjr\_max=2&year\_min=2022).

Performance predicton was considered from both classification and regression type of problem, both methods with advantages and disadvantages, and wide variety of statistical algorithms was used in literature for that prediction (Zhao et al., 2021). Comparison of the performance of the algorithms is also hard due to multisourcity of data (Zhao et al., 2021).

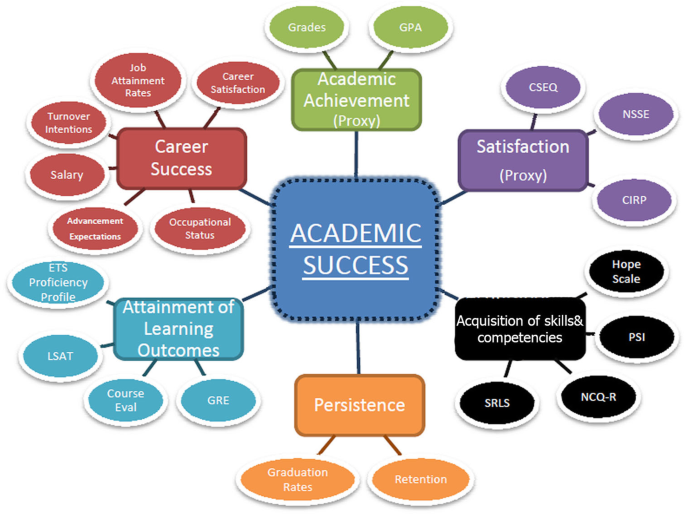


Figure 1 Definition of academic success (York et al., 2015)

One of the most promising tools in this area of prediction is machine learning (ML; 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 reduce bias, offer flexibility, and overall offers a more robust model (Bregant et al., 2025; Hilbert et al., 2021) . This 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 (Figure 2), and more recently even online learning behaviours where open access data is not scarce (Liang et al., 2023) ML models can provide insights that help improve student outcomes.

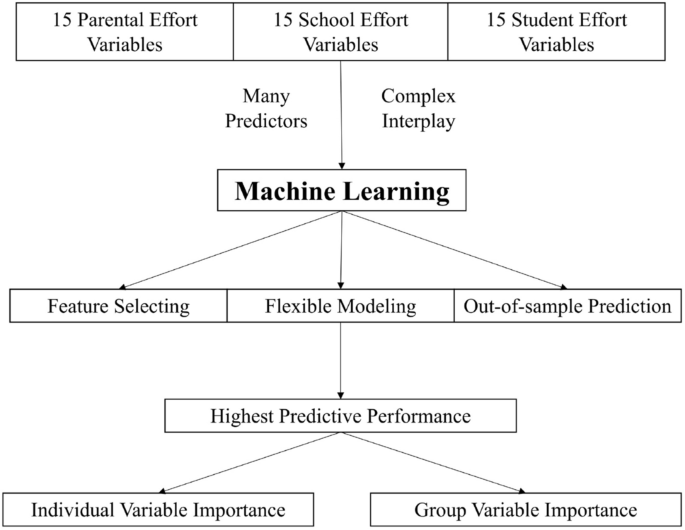


Figure 2: Example of machine learning modelling for student performance prediction by Jin, (2023)

As we have seen, many studies (see Alyahyan & Düştegör (2020) section *Influential factors in predicting academic success*) have explored different factors that contribute to academic success, such as personal effort, family support, and the learning environment. These factors are crucial and have been widely researched, helping schools and educators develop strategies to support students. However, there is a noticeable gap in bibliometric analyses within this field. Understanding the research trends, publication patterns, and thematic developments through a bibliometric approach can provide valuable insights.

Bibliometric analysis is a quantitative method used in systematic literature reviews to assess research productivity and impact (Lim & Kumar, 2024). Many seminal guides were published in this topic e.x. by Donthu et al., (2021); Lim & Kumar, (2024) - with focus on the performance metrics like publications, citations, authorship, trend clustering (i.e. keyword analysis), nomological networking, knowledge gaps, coverage of research, and so on. This technique leverages big data and software tools (e.g., *Gephi*, *OSviewer*, or other already established statistics library tools such as *bibliometrics* in *R*) to provide an objective overview of a research domain .

The present review study is focused on research about machine learning statistics methods for the prediction of academic success of students and therefore opens the following research questions:

RQ1: What are the primary research themes and trends in the domain of academic success prediction using machine learning, based on the analysis of abstracts and keyword co-occurrence?

RQ2: How do publication patterns and cited reference counts vary across different publishers and over time in the research on academic success prediction?

RQ3: What is the distribution of publications over the years, and how has the focus on machine learning in predicting academic success evolved in terms of publication frequency and keyword prominence?

RQ4: Which are the most referenced papers, and how do their cited reference counts correlate with their impact and relevance in the field of academic success prediction?

RQ5: What are the characteristics and distribution of publications by language and publisher city, and how do these characteristics influence the citation patterns and research dissemination?

## Methods

The study was conducted in the academic year 2025 within the subject of *Raziskovalni seminar* on PhD program *Edukacijske vede* on *Faculty of Education* within *University of Primorska*. To establish a rigorous, transparent, reproducible and adaptable review study, the search process was conducted based on the PRISMA protocol.

Preliminary analysis included *SCOPUS, WOS, Taylor & Francis* databases, and register *Digitalni portal Univerze na Primorskem*. A comparison between the databases and registries yielded high overlap for thematic of prediction of academic success using machine learning techniques. For the final bibliometric study *WOS* database was chosen, as it allows science citation index, which was one of the research questions. For the review, articles from all selected databases and registries were accounted for, based on their importance.

For the final report in this article, the final search in the WOS core collection was conducted in January 2025. The search string applied was *("machine learning" OR "ML") AND ("predicting academic success" OR "academic performance prediction" OR "student success prediction") AND ("educational sciences" OR "education")*. As Prisma flowchart indicates in FIGURE, the search results in 286 matches, none of which was removed for the bibliometric study. For the ssystematic literature review (SLR) part, this was later reduced to article publications in the English language which resulted in 283 matches which were further screened based on a title and abstract. From 283 matches the final screening for the review was 12. Following the exclusion criteria of review articles, book chapters, publication year below 2020, WOS category outside education, and articles deemed as not perfectly suitable. We acknowledge that the final cohort of SLR articles is not as big as it should be, following the guidelines of most SLR.

The final selection of 286 articles was analysed using *Python* programming language to construct and visualize the bibliometric network. Despite Python already having a library *pyBibX* designated for bibliometric and scientometric analysis we chose to program our own, which is openly available at https://github.com/borbregant/Doktorat\_all/blob/main/raziskovalni\_seminar/obdelava.ipynb.

Records removed *before screening*:

Records not in English (n = 3)

Records removed for other reasons (n = 86)

Records identified from\*:

Databases (n = 286)

Registers (n = 86) (digitalni portal)

**Identification**

Records screened based on title, abstract and key words

(n = 283)

Records excluded\*\*

(n = 271)

Reports sought for retrieval

(n = 12)

**Screening**

Reports assessed for eligibility

(n = 12)

Reports excluded:

Reason 1 (n = )

Reason 2 (n = )

Reason 3 (n = )

etc.

Studies included in review

(n = 12)

Articles included in bibliographical analysis

(n = 286)

**Included**

Figure 3: PRISMA diagram.

## Results

…

A graph with blue rectangles

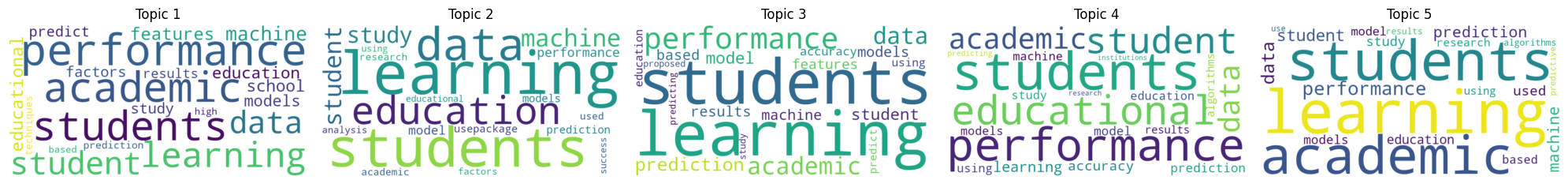
Description automatically generated

A graph of a number of cities

Description automatically generated

A graph of blue bars

Description automatically generated with medium confidence



A diagram of a box plot

Description automatically generated

A graph of blue bars

Description automatically generated with medium confidence

A graph of blue and orange bars

Description automatically generated

## Discussion

…

## Limitations and future directions

…

## References

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

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

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

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

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

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

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

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