

Machine learning on academic education: Bibliometric studies

Hapsari Peni Agustin Tjahyaningtjas^{1*}, *Nanang Husin*², *Hasanuddin Al Habib*³, *Asmunin Asmunin*⁴, *Rindu Puspita Wibawa*⁴, *Alim Sumarno*⁵, *Jesse R Paragas*⁶, and *Endang Susantini*⁷

¹ Department of Electrical Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia.

² Department of Economic and Business, Universitas Negeri Surabaya, Surabaya, Indonesia.

³ Department of Data Sains, Universitas Negeri Surabaya, Surabaya, Indonesia.

⁴ Department of Informatics Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia.

⁵ Department of Educational Technology, Universitas Negeri Surabaya, Surabaya, Indonesia.

⁶ Information Technology Department, College of Engineering, Eastern Visayas State University, Philippines.

⁷ Department of Biology, Universitas Negeri Surabaya, Surabaya, Indonesia.

Abstract. The use of Machine Learning exhibits significant promise in facilitating advancements in the field of education. It is vital to conduct a comprehensive review of existing research to ascertain the significance of utilizing Machine Learning as a viable approach to enhance educational advancements. This bibliometric analysis provides a comprehensive overview of the advancements in the application of machine learning techniques within the field of education. This study utilizes publication and citation data from many academic literature sources to elucidate prominent patterns, areas of research emphasis, and scholarly collaborations within this field. The findings of the bibliometric analysis reveal a significant increase in scholarly attention toward the application of machine learning in the field of education during the past several years. The scope of these investigations encompasses a diverse array of subjects, such as personalized learning, predictive analytics, automated evaluation, learning recommendations, and online exam proctoring. The findings of this study also demonstrate a notable rise in the level of collaboration among scholars from many fields, highlighting the significance of interdisciplinary approaches in tackling the intricate challenges associated with the integration of machine learning in education.

1 Introduction

Machine Learning has advanced rapidly in recent years, allowing machines to learn from examples and experiences, establishing a foothold in the educational sector. On the other hand, Deep learning (DL) is a subfield of ML that focuses on the creation of artificial neural networks (ANN). These networks are comprised of algorithms that are modeled after the structure and function of the human brain [1]. In the area of education, machine

* Corresponding author: hapsaripeni@unesa.ac.id

learning is used worldwide. The application of machine learning in education has resulted in a significant shift in the way we teach and learn by analyzing individual student data and customizing materials and teaching methods to their needs, [2]. This not only allows teachers to give more relevant learning experiences, but it also aids in the prediction of student performance, the identification of factors that influence academic success, and the provision of timely interventions. Furthermore, machine learning is employed for automatic assignment and exam evaluation, saving teachers time and enabling deeper analysis of student progress. As a result, machine learning contributes to the creation of educational settings that are more flexible, effective, and responsive to individual needs [3].

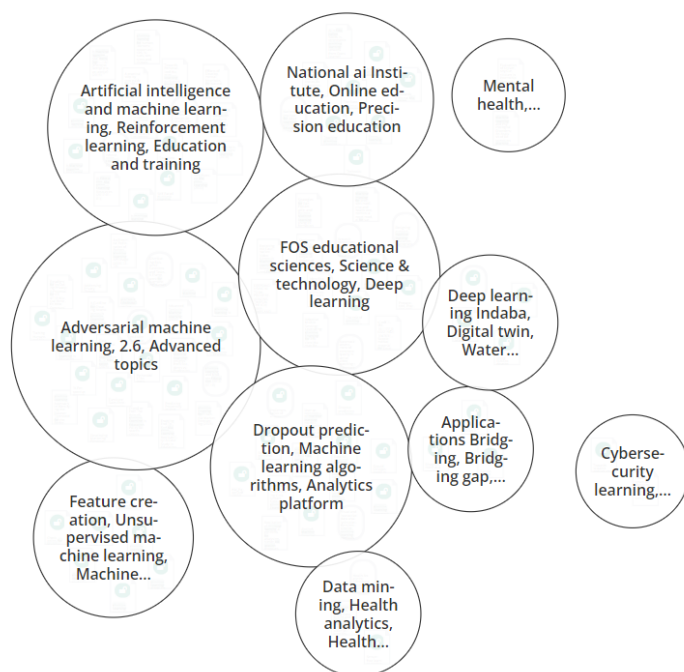


Fig. 1. The area of AI research on education using machine learning

The number of publications, citations, and collaborations between researchers in various disciplines related to education and technology continues to increase. This shows that machine learning has become an integral part of the development of modern education and is the main focus for many researchers who are trying to improve the efficiency and effectiveness of the learning process [4]. It can be seen that the importance of interdisciplinary collaboration in addressing the challenges and opportunities associated with the application of machine learning in educational contexts. For this reason, it is necessary to conduct high research both in quantity and quality and then publish scientifically to encourage discovery and utilization in the field of Machine Learning for Indonesian education. What also needs to be done is to map Artificial Intelligence and Machine Learning research, and then see how far progress related to education is. So, this research answers the following problems: the development of the number of international scientific publications in the field of machine learning from 2015

to 2023; the productivity level of machine learning researchers; and the progress of international publications for machine learning research based on keywords.

The examination of the growth of study in the utilization of Machine Learning holds significant importance for researchers [5]. This analysis is crucial for comprehending the current state of knowledge concerning the art of identifying future research that will be pursued. Through an analysis of the expansion of scholarly inquiry, academics can discern patterns and areas of investigation that necessitate further attention and exploration. In addition to this, researchers can gain insights into the global landscape of Machine Learning research in the education sector, enabling them to assess the relative standing of their research, ascertain its novelty, and identify potential collaborators and available funding opportunities worldwide. In general, monitoring the progression of research concepts is a crucial aspect of remaining informed about the most recent advancements in a certain discipline and propelling the advancement of knowledge through inventive research endeavors.

2 Methods

This study utilized data from international articles on the subject of Machine Learning in education obtained from renowned databases such as ResearchGate, Scopus, and Google Scholar. The data collection process involved utilizing Litmaps and open knowledge maps to search for publications. Machine Learning keywords were then applied to article title categories, abstracts, and keywords within the time frame of 2015 to 2023. However, it was discovered that the research was only documented in 2014. Consequently, data collection was extended from 2014 to 2023, with annual iterations, to ensure greater precision and comprehensiveness in the results. The visualization of development patterns and productivity of authors in the field of machine learning in education can be achieved through the utilization of VOSViewer, a tool commonly employed for such purposes in academic research.

3 Result and discussion

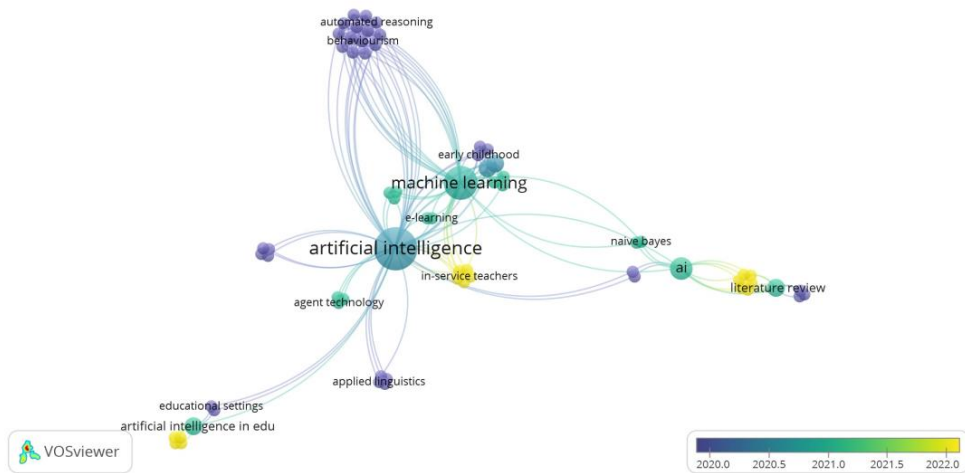


Fig. 2. Overlay visualization map of ML in education

Figure 2 depicts a visual representation of the various study domains where machine learning techniques are being employed within the education field. The visualization was generated by mapping the keywords extracted from the abstracts of 80 articles pertaining to the application of Machine Learning in the field of education, out of a total of 269 articles. The first figure is presented in Figure 2. Based on the analysis of the research landscape spanning from 2020 to 2022, it is evident that the use of Machine Learning exhibits a wide range of applications, encompassing domains such as learning, automated reasoning, applied linguistics, and behavior detection. The visualization reveals a correlation between node brightness and the relative size of the study area associated with each node. Specifically, brighter nodes indicate fewer research areas compared to darker-colored nodes.

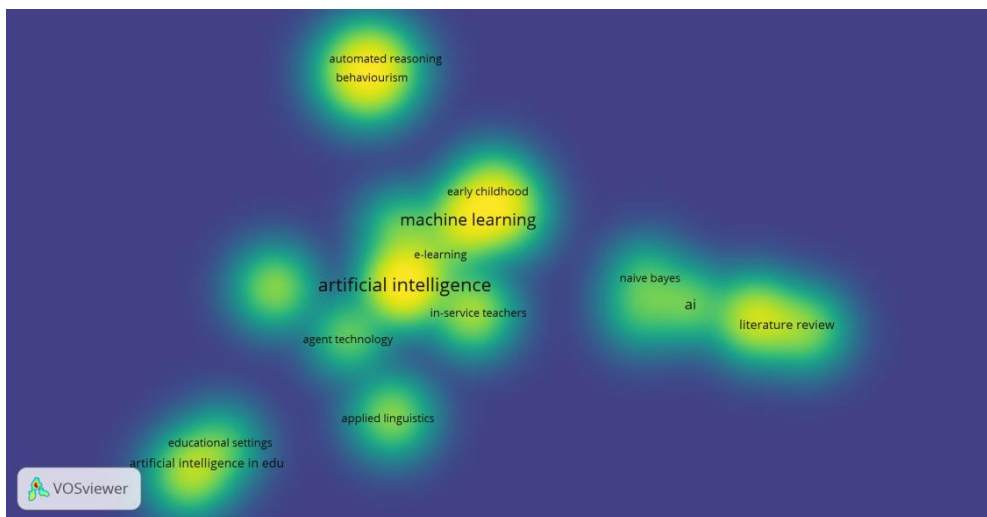


Fig. 3. Density Visualization of the research area of ML in education

Table 1. Number of articles published each year

Year	Number of Articles
2023	66
2022	53
2021	48
2020	23
2019	25
2018	16
2017	15
2016	12
2015	11

Table 2. Number of articles published according to the area of Machine Learning in education

Area of Machine Learning in Education	Number of Articles
Predictive analytic	66
Online exam proctoring	63
Automated evaluation	52
Personalized learning	45
Virtual Assistant	43

Table 1 displays the number of articles that have been acquired from the Scopus and Research Gate databases. The advancement of machine learning (ML) research in the education industry has been progressively growing over time. The 269 papers that were acquired may be classified into four distinct parts, specifically machine learning in the context of personalized learning, predictive analytics, automated evaluation, virtual assistants, and online examinations. According to the data presented in Table 2, it is evident that predictive analytics is the most prevalent area of machine learning research in the education sector, whereas virtual assistants exhibit the lowest level of research activity.

The integration of machine learning into personalized learning has had a profound impact on the educational domain. According to Guilherme (2019) [4], with the utilization of advanced data analysis techniques, machine learning algorithms can comprehend the individual preferences, abilities, and learning styles of each learner. This enables educators to develop customized learning plans that cater to the specific requirements of each student, encompassing the selection of appropriate materials, appropriate levels of challenge, and the implementation of the most efficient instructional strategies. In contemporary educational settings, there has been a shift away from mandating a consistent learning approach for students. Instead, students are now allowed to obtain educational content that is both pertinent and customized to their own

level of comprehension. The integration of machine learning into personalized learning systems offers students an enhanced avenue for realizing their maximum educational capabilities. Simultaneously, this integration empowers teachers to deliver improved assistance and concentrate on addressing the individualized requirements of each student.

The utilization of machine learning in the field of predictive analytics has emerged as a highly potent instrument for enhancing operational efficiency and efficacy across many industries, including the realm of education. Machine learning algorithms are employed within the educational domain to derive valuable insights from a wide range of student data. This enables the anticipation of students' forthcoming academic achievements, the identification of elements that impact their scholastic outcomes, and the detection of prospective hazards, such as students who are susceptible to discontinuing their education [6]. Through the utilization of predictive analysis, educational institutions and educators are empowered to enhance their planning of interventions, provision of specialized support, and development of more efficacious educational programs. Machine learning, as evidenced by studies conducted by Tsiakmaki et al. [7] and Zanirato et al. not only enhances our comprehension of historical educational trends but also enables more accurate and efficient predictions of the future of education [8].

It is a significant advancement when machine learning is used for automated evaluation in a learning environment. Machine learning algorithms possess the capability to efficiently and precisely analyze vast quantities of student responses. According to Sanuvala and Fatima [9], this approach facilitates the effective evaluation of examinations and assignments, minimizing the time demands on teachers or tutors. Furthermore, it has been observed that machine learning algorithms can detect and analyze patterns and trends within students' responses, hence aiding educators in gaining insights into certain areas that may necessitate targeted instructional interventions [10]. Automated evaluation further facilitates the maintenance of consistency in scoring, a crucial aspect in upholding fairness and objectivity throughout the scoring procedure. The integration of machine learning into automated assessment systems enhances the efficiency of education, facilitates prompt feedback delivery to students, and enables teachers to dedicate their attention to the pedagogical and innovative dimensions of their instructional practices.

Machine learning algorithms are utilized to examine student data, encompassing their learning history, preferences, level of comprehension, and academic advancement [2]. Based on the provided information, the system possesses the capability to suggest supplementary educational materials, textbooks, online resources, or even courses that align with the student's interests and aptitudes [11]. This not only facilitates pupils in attaining a more pertinent and captivating learning encounter but also enhances the overall caliber of education. Learning recommendations derived from machine learning algorithms can effectively enhance student motivation by offering them suitable tasks that align with their individual ability levels. Therefore, this particular strategy possesses the capacity to revolutionize how pupils obtain and undergo the process of learning, thereby enhancing its level of interactivity and efficacy.

The application of machine learning in the context of virtual assistants has introduced novel approaches within the field of education. According to a study conducted by S.

Iqbal et al. [12], machine learning-based virtual assistants have the capability to engage in direct interactions with students, discern their requirements, and offer suitable recommendations for their learning endeavors. Singh and Jaiswal [13] assert that the utilization of machine learning techniques enables virtual assistants to enhance their recommendations iteratively by leveraging interactions with students and monitoring their learning progress. This not only enhances the individualization of the learning process, but also provides students with supplementary resources, reading materials, exercises, or materials that are pertinent to a specific subject matter. This particular form of virtual assistant can aid students in answering inquiries, offering clarifications, and providing personalized assistance, hence enhancing the accessibility and interactivity of the learning experience. According to Almusaed et al. [14], the utilization of machine learning techniques enhances the cognitive capabilities of this virtual assistant, hence facilitating a more optimized and proficient learning encounter for students.

A crucial method for preserving the integrity of online assessments is the use of machine learning. According to Kaddoura and Gumaei [15], the utilization of machine learning algorithms enables the surveillance of student activities during examinations, facilitating the identification of instances of academic dishonesty. These algorithms are capable of issuing alerts in response to potentially suspect behaviors, including but not limited to screen sharing, accessing unauthorized resources, or engaging in communication with other individuals. The utilization of this technology has the potential to enhance the efficacy and automation of online exam proctoring, hence alleviating the workload of human proctors and offering enhanced safeguards against academic dishonesty [16]. The utilization of machine learning in exam proctoring necessitates ethical and privacy deliberations, as well as the need for transparency to uphold fairness and foster student confidence in the assessment procedure.

4 Conclusion

A bibliometric analysis was conducted to examine the use of Machine Learning in the realm of education, to present a comprehensive perspective to academics of the growing body of research in this domain. The study of the 269 articles revealed that there was a significant focus on the application of machine learning (ML) in the domain of predictive analysis, indicating a high level of interest in this area. Conversely, the research pertaining to virtual assistants exhibited comparatively lower levels of interest among scholars and practitioners.

Acknowledgments

This study is a part of the Innovation Product Development Research Project, which is financially supported by the Directorate of Innovation Rankings and Scientific Publications, administered by the Institute for Research and Community Service Universitas Negeri Surabaya Indonesia.

References

1. J. Kufel, et al., *Diagnostics* **13**, 15 (2023)
2. K. Seo, J. Tang, I. Roll, S. Fels, D. Yoon, *International Journal of Educational Technology in Higher Education* **18**, 1 (2021)
3. E. M. Onyema et al., *Comput Intell Neurosci.* **2022** (2022)
4. A. Guilherme, *AI Soc.* **34**, 1(2019)
5. J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, A. Durán-Domínguez, *Applied Sciences* **10**, 3 (2020)
6. M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, *Comput Intell Neurosci* **2018** (2018)
7. M. Tsiakmaki, G. Kostopoulos, G. Koutsonikos, C. Pierrakeas, S. Kotsiantis, O. Ragos, *Predicting University Students' Grades Based on Previous Academic Achievements*, in 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), Zakynthos, Greece: IEEE Xplore (2018)
8. S. Zanirato, Maia, et al., *World* **2023** **4** 2 (2023)
9. G. Sanuvala, S. S. Fatima, *A Study of Automated Evaluation of Student's Examination Paper using Machine Learning Techniques*, in Proceedings - IEEE 2021 International Conference on Computing, Communication, and Intelligent Systems, ICC CIS, (2021)
10. A. Kashi, S. Shastri, A. R. Deshpande, J. Doreswamy, G. Srinivasa, *A Score Recommendation System Towards Automating Assessment In Professional Courses*, Proceedings - IEEE 8th International Conference on Technology for Education, T4E 2016, (2017)
11. Y. Yamasari, R. A. Ahmad, H. P. A. Tjahyaningtijas, A. Qoiriah, N. Rochmawati, A. Prihanto, *Improving the Performance of Classification via Clustering on the Students' Academic Performance using Stacking Algorithm*, in Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR) (2023)
12. M. Z. Iqbal, A. G. Campbell, *Touchless Hand Interaction and Machine Learning Agents to Support Kinesthetic Learning in Augmented Reality*, in SIGCSE 2021 - Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (2021)
13. N. Singh, U. C. Jaiswal, *Analysis of Student Study of Virtual Learning Using Machine Learning Techniques*, *International Journal of Software Science and Computational Intelligence*, IJSSCI (2022)
14. A. Almusaed, A. Almssad, I. Yitmen, R. Z. Homod, *Education Sciences* **13**, 7(2023)
15. S. Kaddoura, A. Gumaei, *Intelligent Systems with Applications* **16** (2022)
16. D. R. E. Cotton, P. A. Cotton, J. R. Shipway, *Innovations in Education and Teaching International* **60** (2023)