Issue VI, 22 November 2023
e-ISSN 2707-9481 Institute of Metallurgy and Ore Beneficiation, Satbayev University, Almaty, Kazakhstan
ISBN 978-601-323-356-7 https://doi.org/10.31643/2023.34

Zulfa Safina Ibrahim
Yogyakarta State University
E-mail: zulfasafina.2022@student.uny.ac.id

Jumriani Sultan
Yogyakarta State University
E-mail: jumrianisultan.2022@student.uny.ac.id

Fina Rifqiyah
Yogyakarta State University
E-mail: finarifqiyah.2022@student.uny.ac.id

Heri Retnawati
Yogyakarta State University
E-mail: heri_retnawati@uny.ac.id

A Visualized Bibliometric Analysis for Mapping Research Trends of Machine Learning in Academic Research

Abstract: This study aims to conduct a comprehensive bibliometric analysis of machine learning research within the academic research from 2013 to 2023. The goals include comprehending how publication trends have changed over time, pinpointing important fields of interest, and clarifying how machine learning and scholarly research interact. This study's methodology combines a bibliometric framework with a qualitative descriptive approach. The data were collected from Scopus database, which consist of 4.655 documents for analysis. The analysis is executed using R studio and Biblioshiny. The result shows that the rise of machine learning in academic research between 2013 and 2023 suggests that this field of study will continues to be a trending area for academic exploration.

Keywords: Bibliometric Analysis, Machine Learning, Academic Research, Machine Learning in Education.

Cite this article as: Ibrahim Z. S., Rifqiyah F., Sultan J., Retnawati H. (2023). A visualized Bibliometric Analysis for Mapping Research trends of Machine Learning in Academic Research. *Challenges of Science*. Issue VI, 2023, pp. 271-279. https://doi.org/10.31643/2023.34

Introduction

Evaluation of student achievement is the main focus in the world of education, especially in assessing student achievement in educational institutions. Educational institutions operate in a highly competitive and demanding environment. Most schools face difficulties in providing high-quality education, evaluating student performance, and understanding the future demands of their learners (Issah et al., 2023). Technological developments and information-based decision making have transformed our world over the past decades, bringing about the birth of the knowledge economy (Maphosa & Maphosa, 2023). One of them is the development of Artificial Intelligence itself (Zemel et al. 2013; Rohde et al., 2023). Advances in technology and data availability have enabled rapid developments in various aspects of machine learning, ranging from understanding basic concepts to implementing highly complex models. Along with these developments, it is important to understand academic research trends in the machine learning domain.

Machine learning has a crucial role in the exploration of educational data, enabling the ability to make predictions in the education sector. One of the advantages of this approach is its ability to identify questions that frequently arise (Pallathadka et al., 2023). Gaurav & Patel (2020) state, machine learning is a scientific discipline that aims to develop computer algorithms to turn data into intelligent actions. In recent decades, enormous growth in data and computing power has driven the development of statistical methods for analyzing complex data. Machine learning algorithms are needed to make machines understand and respond to data, with simulation of human learning processes, reasoning from incomplete data, exploration of new discoveries, and analysis of data trends being part of the process (Wu & Zheng, 2021).

The use of machine learning in higher education is experiencing significant growth. Some examples of its implementation in the world of education include the development of systems that can project student retention rates and provide students with feedback tailored to individual needs (Tahiru et al., 2023). In recent years, machine learning methods have been used to analyze student data, which is in line with efforts to improve information processing through data mining (Cardona et al., 2019a). For example, Machine Learning has been successfully applied to predict students who are at high risk of dropping out of school. Therefore, machine learning is very important in the world of education.

Bibliometric analysis is a very relevant method for the research questions in this study because it can highlight the quality of publications in a particular field (Tahiru et al., 2023). Bibliometric analysis has emerged as a highly effective tool in mapping and analyzing research trends in various scientific fields. In the context of machine learning, bibliometric analysis can provide valuable insights into the development and dynamics of academic research in this domain. This allows us to identify current research, influential authors, key concepts, and emerging topic trends.

Research using bibliometric analysis to describe the use of machine learning in higher education is limited. This could create further research opportunities to provide insight into future trends in higher education (Tahiru et al., 2023). The purpose of this article is to review research that has been conducted on machine learning-based academic research from 2013 to 2023. This article provides an in-depth bibliometric review that examines patterns and trends regarding key publications, country contributions, significant topics, and trends in the field this research. This study focused on answering the following research questions:

Q1. How has the publication output of machine learning research in academic research evolved based on bibliometric analysis of Scopus data from 2013 to 2023?
Q2. Which areas exhibit the greatest interest in research related to machine learning in the academic sphere over the last decade, as assessed through bibliometric analysis of the Scopus database?
Q3. Which areas exhibit the greatest interest in research related to machine learning in the academic sphere over the last decade, as assessed through bibliometric analysis of the Scopus database?

Research Methods

This research uses descriptive qualitative research with a bibliometric approach. Bibliometric analysis is a structural analysis of research results, and groups information obtained from metadata in the form of articles, books, conferences, reviews (Sri Rahayu & Sofian Hadi, 2023). The data used was sourced from Scopus with a period of 2013-2023 with the keywords "Machine Learning" and "Academic research" so that data was obtained for 4655 documents. The data obtained was analyzed using R studio and Biblioshiny. The following is the flow of the research

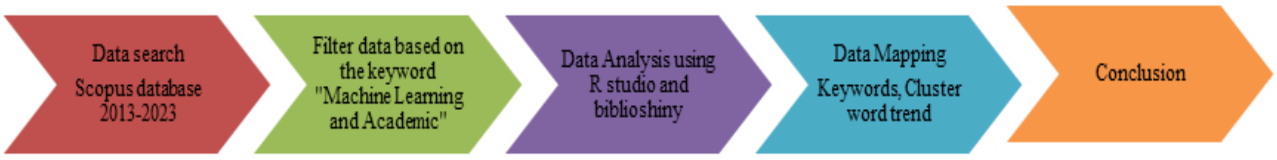

Fig 1. Flow Research

Literature search and Data collection

In total, 4655 papers were collected after refining based on the inclusion and exclusion criteria shown in Table 1. These data were exported for analysis. The search string for this study is made up of several keywords joining together "Machine Learning" AND "Academic". The search and retrieval of the data were conducted on November 5th, 2023.

Table 1. Inclusion and exclusion criteria for retrieving the dataset

	Code	Criteria	Data Extraction
Inclusion Criteria	IC1	Articles containing one of the keywords "Machine Learning" and "Academic"	6255
	IC2	Published articles between 2013 to 2023	5847
	IC4	Articles in journals and conferences papers	4864
	IC5	Documents written in the English language	4783
Exclusion Criteria	EC1	Articles with publication stage "in press."	128
		Total Articles	**4655**

Research Experiment

Results and discussion of findings are presented in this section to reflect (i) Data information and Overall Publications Growth; (ii) Countries and research institutions distributions; (iii) Most Relevant Keywords of Machine Learning in Academic Research.

Data information and Overall Publications Growth. In this section, this study will focus on scientific productions in the field of machine learning in academic research from 2013 to 2023. Analyzing annual scientific production helps researchers, institutions, and policymakers understand the trends and priorities in a particular field (Dou, 1985). The annual scientific production of articles is presented in Table 1. The data reveals a significant increase in the annual scientific production of articles in the field of machine learning over the past decade. In 2013, there are 42 articles which had been published within this field, and the number of publications increased steadily each year (see Fig. 2). By 2022, the production of the articles in this field reached an impressive number, with 1.124 documents, signifying a more than twenty-fold increase over the ten-year period. This substantial growth reflects the increasing attention, engagement, and investment in machine learning research among scholars.

Alongside the growth in the number of articles, the mean total citations per year per article exhibit dynamic patterns (see Table 2). In 2018, despite the significant increase in the number of articles (267), the mean total citations per year per article also showed a considerable surge, reaching 5.95. This suggests that the research output of that year had a remarkable impact and garnered substantial attention within the academic community. However, in the subsequent years, such as 2022 and 2023, while the number of articles remained high (1,124 and 989, respectively), the mean total citations per year per article declined to 1.96 and 1.04, respectively. This fluctuation indicates variations in the visibility and impact of research output, which is probably influenced by the evolving nature of the field and competition for attention.

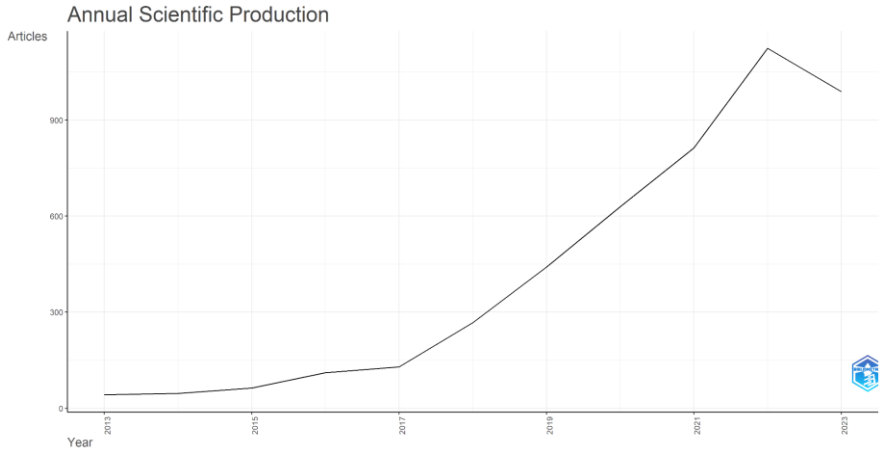

Fig 2. Annual scientific production over the years

Table 2. Articles production and average citation per year

Year	Articles	Average Citation PerYear
2013	42	2,89
2014	46	3,26
2015	63	3,64
2016	111	2,46
2017	129	3,49
2018	267	5,95
2019	441	3,82
2020	630	4,34
2021	813	3,27
2022	1124	1,96
2023	989	1,04

Countries and research institutions distributions

Distribution of articles by country. Every year, many countries conduct research in the field of differentiated instruction. However, when we examine the productivity of these countries, the top ten countries are their countries of origin. When research published by Scopus is displayed or presented based on the country conducting the research, the results can be seen in the chart in Figure 3.

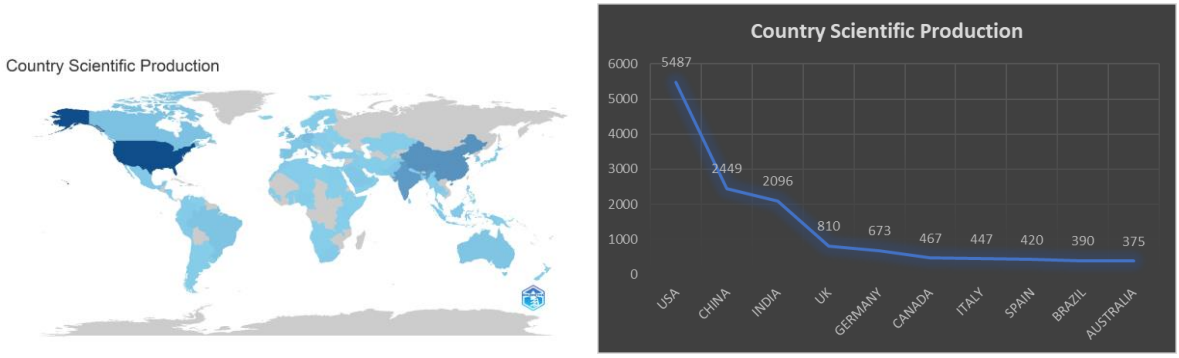

Fig 3. Distribution of articles by country

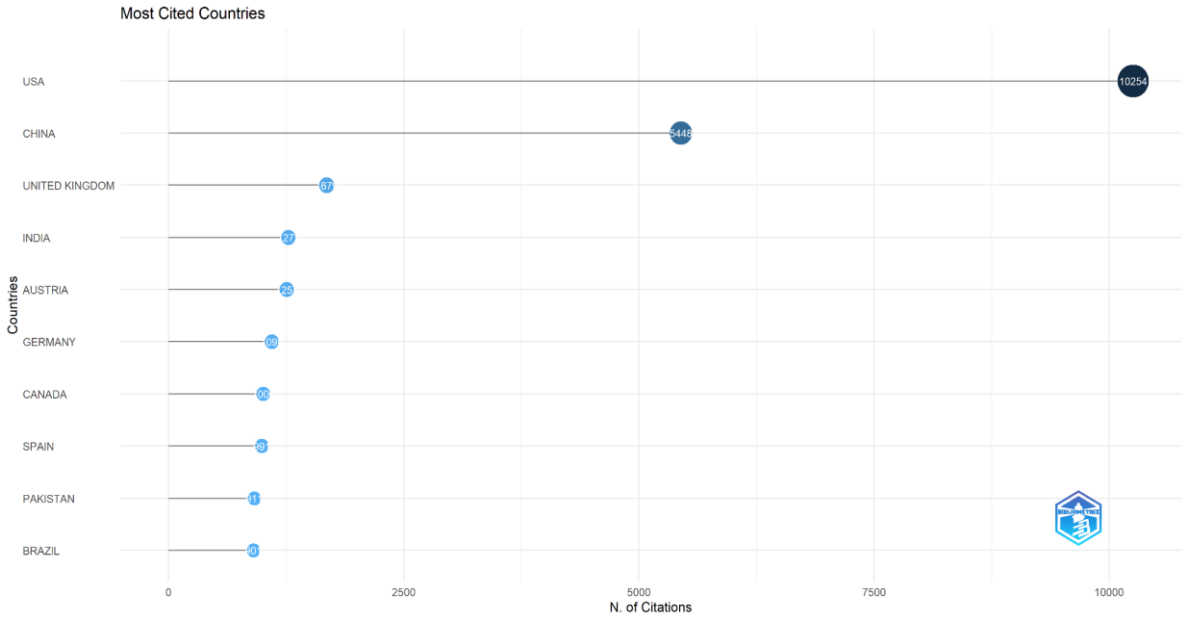

Fig 4. Most citation country

Providing data on the most productive countries in generating articles on machine learning in academic research. The data indicates that the USA is the most productive, with 5487 articles related to this topic. China is in the second position with 2449 articles, followed by India, the UK, and Germany as the top five most productive countries. This suggests that the USA holds a dominant position in machine learning research in academic studies when compared to other countries visible in Figure 4. The USA's most productive position may be influenced by several factors, such as a high number of universities and research institutions, research funding availability, and the early adoption of machine learning approaches in the country.

***Most citation country*.** The following provides information about the top 10 countries with the highest number of citations obtained from various sources and authors. Figure 4 presents the results of the bibliometric analysis. Figure 4 displays information about the countries and their highest number of citations tracked by Scopus. Based on the data in Figure 4 articles from the United States have received the highest number of citations (10,254) among all articles on the topic of machine learning in academic research. This is followed by articles from China, the United Kingdom, India, and Australia, which are among the most cited articles on this topic. The fact that the United States has the highest number of citations indicates that their research has had a significant impact on the field of machine learning in academic research. However, it should be noted that a high number of citations does not always reflect research quality, as other factors such as topic popularity or research availability can influence it. The presence of other countries on the list of the most cited articles in human resource management also suggests that their research has made substantial contributions to the field. Overall, the data presented in Figure 4 offers valuable information for researchers interested in machine learning in academic research, helping them identify key authors and articles in the field. Additionally, it can provide insights into research impact in the field and future research directions.

Most productive affiliate. The following information presents the most productive affiliations in the field of machine learning within academic research. The outcomes of the bibliometric analysis are depicted in Figure 5.

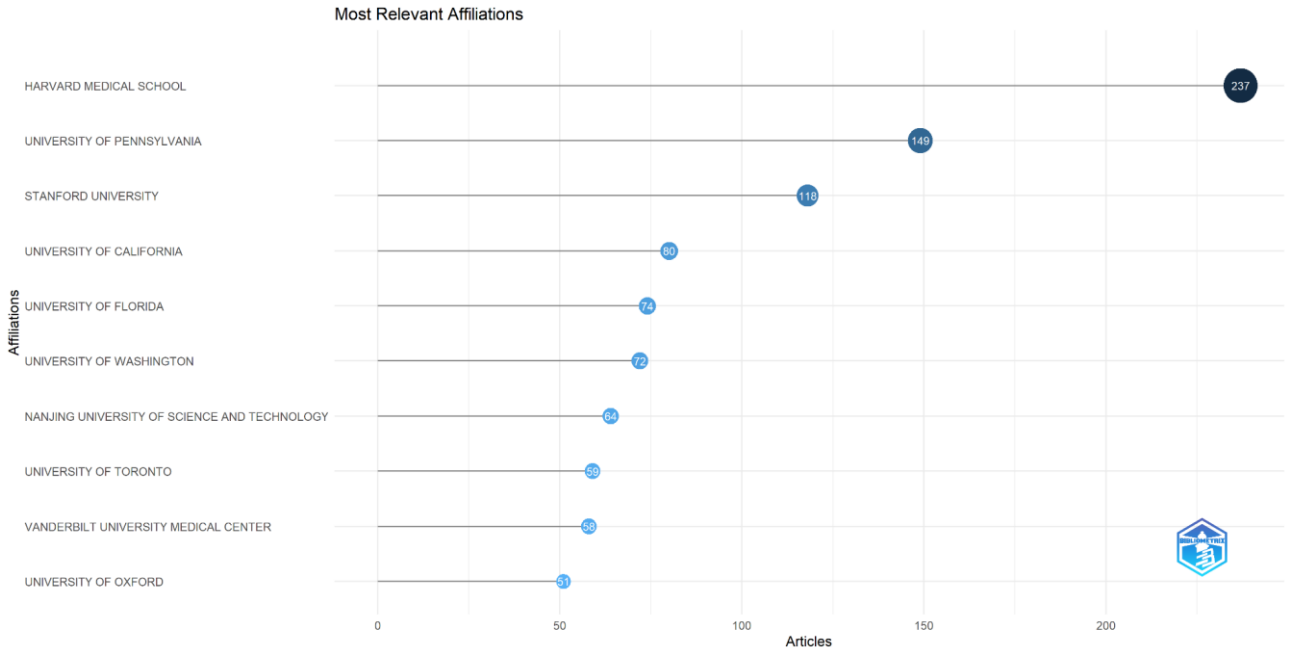

Fig 5. Most productive affiliate

As depicted in Figure 5, Harvard Medical School stands out as the leading institution in terms of publishing articles in the field of machine learning within academic research. With 237 articles to its name, it surpasses other institutions by a significant margin. This highlights Harvard Medical School's considerable involvement in this area, signifying its substantial contribution to advancing the knowledge of machine learning in academic research. Furthermore, the University of Pennsylvania also demonstrates notable activity in publishing articles on this subject, having produced 149 articles related to machine learning in academic research and securing the second position among affiliations with the most publications. In addition to these prominent affiliations, three others—Stanford University, the University of California, and the University of

Florida—also feature in the top five affiliations with the highest publication output in the realm of machine learning within academic research.

This indicates the widespread interest of researchers from various universities around the world in the field of machine learning within academic research. This data serves as a valuable resource for researchers and academics, enabling them to stay informed about the latest developments in the field of machine learning in academic research and facilitating the identification of potential research partners or collaborators in relevant universities. Furthermore, it offers insights into the global landscape of research and intellectual contributions in the field of machine learning within academic research. By identifying the most active affiliations in publishing within this field, researchers can make informed decisions about the direction of their research to further enrich the understanding of machine learning in academic research.

Table 3. Most productive sources

Sources	Articles
LECTURE NOTES IN COMPUTER SCIENCE (INCLUDING SUBSERIES LECTURE NOTES IN ARTIFICIAL INTELLIGENCE AND LECTURE NOTES IN BIOINFORMATICS)	130
ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES	100
LECTURE NOTES IN NETWORKS AND SYSTEMS	84
IEEE ACCESS	81
COMMUNICATIONS IN COMPUTER AND INFORMATION SCIENCE	69
PLOS ONE	56
ADVANCES IN INTELLIGENT SYSTEMS AND COMPUTING	53
APPLIED SCIENCES (SWITZERLAND)	52
CEUR WORKSHOP PROCEEDINGS	46
SUSTAINABILITY (SWITZERLAND)	38

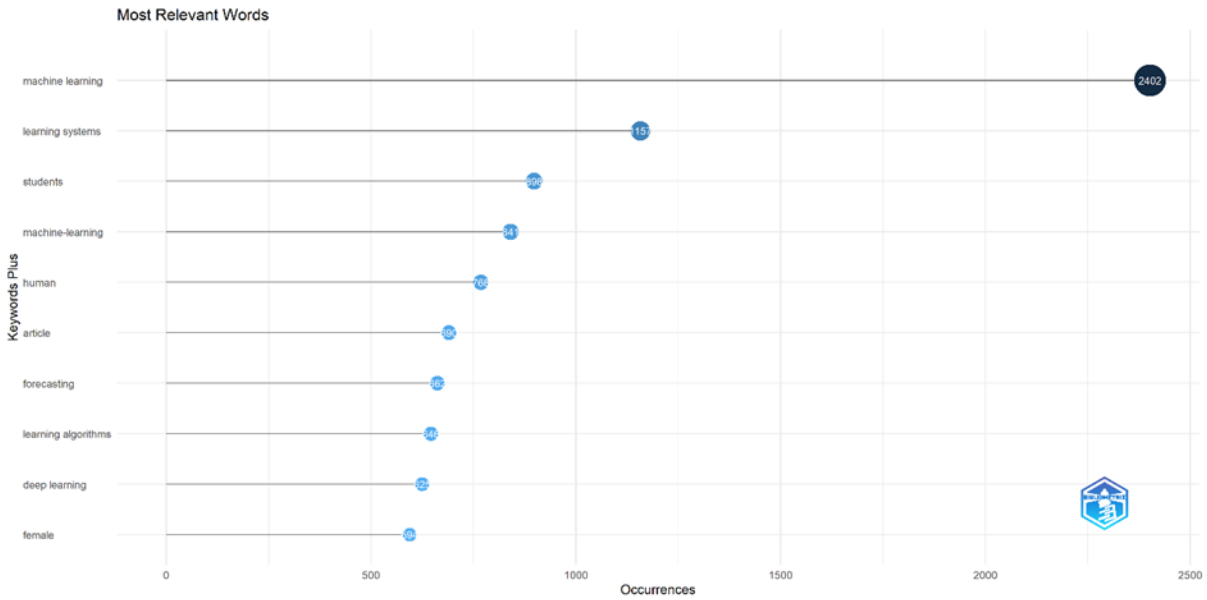

Fig 6. Most Relevant Words

***The most productive sources*.** Table 3 provides information about the most productive journals and publication sources in terms of publications related to machine learning in academic research. From the table, "LECTURE NOTES IN COMPUTER SCIENCE (INCLUDING SUBSERIES LECTURE NOTES IN ARTIFICIAL INTELLIGENCE AND LECTURE NOTES IN BIOINFORMATICS)" is the most productive publication with 130 articles. It is followed by "ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES" with 100 articles. Next, there is "LECTURE NOTES IN NETWORKS AND SYSTEMS" with 84 articles, "IEEE ACCESS" with 81 articles, and "COMMUNICATIONS IN COMPUTER AND INFORMATION SCIENCE" with 69 articles as the top five in the number of published articles. This table also provides important information about the most productive and influential publication sources in the field of machine learning in academic research.

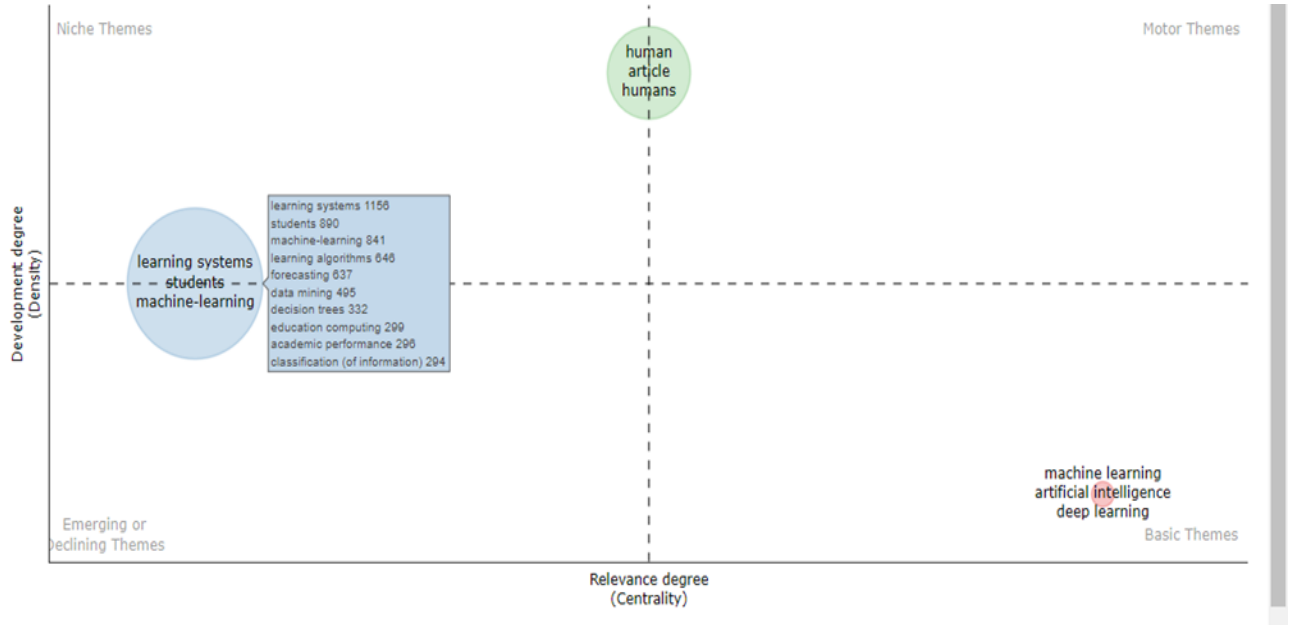

Fig 7. Thematic Map

***Most Relevant Keywords of Machine Learning in Academic Research*.** The thematic map contains three clusters by binding to the author's keywords, with the variables author's name, publication title, keywords and number of citations. Special topics in the field of machine learning include human-computer interaction. Basic themes can be classified into learning systems, students, machine learning, human, artificial intelligence and deep learning. In the thematic map, a bibliographic index of the top three Machine Learning, Learning Systems and Students is obtained.

Fig 8. Word Map

The Word Map illustrates the relationship of the Machine Learning topic to other topics and depicts related research trends. In addition, it provides future research opportunities. The brighter and larger the font size of the keyword indicates that the topic is widely studied or of interest, such as "Machine Learning, and both learning systems and is followed by students with each document index like figure 7 of 2402, 1157 and 898. The smaller the font size of the keyword This means that the discussion regarding this topic is still not well studied so this could be a candidate for discussion that can be discussed in the future.

Research Discussions

Annual scientific publication. The data reveals a noteworthy trajectory in the annual scientific production of machine learning articles. From 2013 to 2022, there was a consistent upward trend in publication output, demonstrating the field's sustained growth and relevance (see Fig. 1). However, in 2022 and 2023, there is a notable shift, with a decrease in both the number of articles and the mean total citations per year per article. The correlation between publication growth and citation patterns highlights the need to consider both the quantity and quality of research output in understanding the dynamics of the field. While increased publication growth is a positive indicator of research activity, the impact and visibility of individual publications may vary, reflecting the evolving nature of the field and the competition for scholarly attention.

***Countries and research institutions distributions*.** In the context of academic research on machine learning, Figure 7 displays the categorization of relevant topics based on current research trends. In the first group, there are topics that are currently emerging in research trends, such as learning systems, students, learning algorithms, prediction, data mining, decision trees, educational computing, academic performance, and classification. In the second group, there are topics that are still under research scrutiny, related to humans and articles. Meanwhile, the third group comprises well-established and relevant topics in the context of machine learning in academic research, including machine learning itself, artificial intelligence, and deep learning. There are also topics situated between the first and second groups, indicating that these topics continue to be primary focuses in research and development of machine learning in an academic context.

Figure 7 and 8 provides valuable information for researchers in selecting research topics that align with current research trends and have the potential to contribute to the improvement of education. In bibliometric analysis, there is a method called conceptual structure mapping used to identify and analyze the relationships between concepts in academic literature. This method involves keyword mapping and cluster analysis of scientific documents in relevant databases. It helps researchers understand research trends and identify research opportunities that need further exploration.

Most Relevant Keywords of Machine Learning in Academic Research. *Bibliometric* analysis in machine learning in the academic field has a strong relationship. This can be seen in Figures 2 and 3 which indicate that machine learning and system learning in the last 10 years have become a highly discussed topic in the world, especially education. This is also supported by the opinion of Song & Wang (2020) that machine learning influences students' metacognition, emotions, motivation and learning skills. Machine learning has contributed to the academic field, especially the development of student-centered learning, increasing motivation and response or activeness in learning (Okagbue et al., 2023).

Thematic analysis shows a combination of bibliography and keywords of interest which are called clusters. Thematic clusters include; machine learning, learning systems and students. This finding is similar to the findings of (Akdeniz & Ozdine, 2021) who found that the terms "machine learning" and "student" influence a person's academics by changing the way of learning and better ways of teaching. Apart from that, the increase in research on machine learning from 2013 to 2023 has increased significantly, meaning that this is still a trend to be researched from various perspectives to provide a better academic impact.

Conclusions

This comprehensive bibliometric analysis of machine learning research within academic research from 2013 to 2023 yields valuable insights into the evolving landscape of machine learning within the educational context. The analysis covers multiple facets, including annual scientific publication trends, the categorization of research topics, and the thematic clusters that reflect the most relevant keywords in machine learning in an

academic setting. The results show the trajectory of annual scientific production in machine learning articles reveals a compelling narrative. From 2013 to 2022, there is a consistent upward trend, indicating the sustained growth and relevance of the field. The data underscores the field's sustained growth, emphasizes the importance of considering both quantity and quality in research assessment, and guides researchers in identifying relevant and impactful research topics. The thematic analysis highlights the enduring influence of machine learning and students in academia, paving the way for further exploration to enhance educational practices.

For future research, it is recommended to emphasize the quality of machine learning research in academic contexts, exploring emerging research topics in machine learning within academia and their influence on educational practices can guide the development of innovative pedagogical approaches. It also can enrich the understanding of machine learning's role in shaping the future of education.

Cite this article as: Ibrahim Z. S., Rifqiyah F., Sultan J., Retnawati H. (2023). A visualized Bibliometric Analysis for Mapping Research trends of Machine Learning in Academic Research. *Challenges of Science*. Issue VI, 2023, pp. 271-279. https://doi.org/10.31643/2023.34

References

Cardona, T., Cudney, E., & Snyder, J. (2019a). Predicting degree completion through data mining [Paper presentation]. Proceedings of the ASEE Annual Conference & Exposition, Tampa, FL, United States.

Dou, H., & Hassanaly, P. (1985). Coal liquefaction case study: impact, potential, forecast. Online Information Review, 9, 137-146. https://doi.org/10.1108/EB024179.

Gaurav, K. A., & Patel, L. (2020). Machine Learning With R. https://doi.org/10.4018/978-1-7998-2718-4.ch015

Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. Decision Analytics Journal, 7(March), 100204. https://doi.org/10.1016/j.dajour.2023.100204

Maphosa, V., & Maphosa, M. (2023). Artificial intelligence in higher education : a bibliometric analysis and topic modeling approach. Applied Artificial Intelligence, 37(01). https://doi.org/10.1080/08839514.2023.2261730

Okagbue, E. F., Ezeachikulo, U. P., Akintunde, T. Y., Tsakuwa, M. B., Ilokanulo, S. N., Obiasoanya, K. M., Ilodibe, C. E., & Ouattara, C. A. T. (2023). A comprehensive overview of artificial intelligence and machine learning in education pedagogy: 21 Years (2000–2021) of research indexed in the scopus database. Social Sciences & Humanities Open, 8(1), 100655. https://doi.org/10.1016/j.ssaho.2023.100655

Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. Materials Today: Proceedings, 80(July), 3782–3785. https://doi.org/10.1016/j.matpr.2021.07.382

Rohde, N., Flindt, N., Rietz, C., & K. Kassymova, G. (2023). How e-learning programs can be more individualized with artificial intelligence – a theoretical approach from a pedagogical point of view. Muallim Journal of Social Sciences and Humanities, 7(3), 1-17. https://doi.org/10.33306/mjssh/240

Song, P., & Wang, X. (2020). A bibliometric analysis of worldwide educational artificial intelligence research development in recent twenty years. Asia Pacific Education Review, 21(3), 473–486. https://doi.org/10.1007/s12564-020-09640-2

Sri Rahayu, N., & Sofian Hadi, M. (2023). Pembelajaran berdiferensiasi dan berpikir kreatif matematis: Penggunaan Vos Viewer dalam Pemetaan Literatur Analisis Bibliometrik. Union: Jurnal Ilmiah Pendidikan Matematika, 11(1), 35–40. https://doi.org/10.30738/union.v11i1.14270

Tahiru, F., Parbanath, S., & Agbesi, S. (2023). Machine Learning-based Predictive Systems in Higher Education: A Bibliometric Analysis. Journal of Scientometric Research, 12(2), 436–447. https://doi.org/10.5530/jscires.12.2.040

Wu, B., & Zheng, C. (2021). An Analysis of the Effectiveness of Machine Learning Theory in the Evolution of Education and Teaching. Hindawi Journal, Northeast Normal University (China), October 11, 2021, pp. 1-10.

Zemel, R., Y. Wu, K. Swersky, T. Pitassi, and C. Dwork (2013). Learning fair representations. International Conference on Machine Learning. 28, pp. 325–33. Atlanta: PMLR.