

The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2022

Volume 25, Pages 17-25

IConSoS 2022: International Conference on Social Science Studies

Bayesian Network Approach in Education: A Bibliometric Review Using R-Tool and Future Research Directions

Maran CHANTHIRAN

Universiti Pendidikan Sultan Idris

Abu Bakar IBRAHIM

Universiti Pendidikan Sultan Idris

Mohd Hishamuddin ABDUL RAHMAN

Universiti Pendidikan Sultan Idris

Punithavili MARIAPPAN

Universiti Pendidikan Sultan Idris

Abstract: The development and multiple variations in technology and science have endured the education. Nevertheless, education is one of the primary components that uphold the development of a country. In the meantime, diverse technologies have been introduced to blend in education. For example, Bayesian Networks is a probability-based data modelling approach that illustrates a set of variables and their conditional dependencies through a Directed Acyclic Graph (DAG). Each node formed inside the graph has a Conditional Probability Table (CPT). Therefore, the endurance of this bibliometric review is to identify peer-reviewed literature on the Bayesian network approach in education. Scopus citation databases are used in the data-gathering phase. In addition, PICOS Framework and PRISMA approach were obtained and analysed for keyword search on the research topic. This bibliographic data of articles published in the journals over ten years were extracted. R-tool and VOS viewer were used to analyse the data contained in all journals and articles. This bibliometric review shows the usage of the Bayesian network approach in education, especially in educational application development. The findings from 87 articles extracted show that teaching and learning activity delivery and educational management have improved. The findings show an increasing trend in published studies related to the Bayesian network in education. Next, the United Kingdom and the United States became highly productive countries in the publication of studies within the scope of the Bayesian network. Next, interdisciplinary became the primary choice in the publication of studies in the field of Bayesian networks. The level of predictive accuracy generated through the Bayesian network approach improves the quality of educational application development. However, the findings of previous studies indicate that there is a need to extend the Bayesian network approach in education.

Keywords: Bayesian Network, R-tool, Bibliometric Analysis, Technology, Teaching Application

Introduction

Education has undergone drastic developments since the world faced the Covid-19 pandemic. 80% of schools had been affected by the Covid-19 pandemic (Azrin & Nurfaradilla, 2021). The Covid-19 pandemic has drastically impacted the world education system (Teras et al., 2020). Practices in teaching and learning activities that have long depended on the classroom are also affected, and all teaching activities are carried out with the help of technology (Chabibie, 2020).

- This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 Unported License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

© 2022 Published by ISRES Publishing: www.isres.org

Based on a United Nations Educational, Scientific and Cultural Organization (UNESCO) (2019) report, as many as 74% of students worldwide are affected by this Covid-19 pandemic. This situation makes teachers and students rely on technology solely to collaborate to access Education. The post-Covid-19 situation, has applied the importance of technology in the world of Education. As a result, there is a need to provide learning aids that students can access even as the world faces a pandemic such as Covid-19. According to Raja and Nagasubramani (2018), technology has become compulsory for fulfilling 21st-century education skills among students. According to Al-Qozani and Aleryani (2018), the use of technology has opened up space in the world of Education to facilitate students' access to Education from a place without time limitations and gaps between teachers.

Developments in technology today recreate a vital role in life. The 21st-century era is often considered the era of technology (Raja & Nagasubramani, 2018). It is also seen as the basis of economic growth of a country where the effect can be felt in every field, one of which may be Education. The Covid-19 pandemic has prompted the ranks of teachers to explore e-teaching methods and platforms such as Google Classroom, Google Meet, and Zoom. Cai et al., (2020) stated that teachers played a critical role in adapting e-teaching during the Covid-19 pandemic period, which had suddenly changed the pattern of teaching delivery method. As a result, the quality of the teaching and learning process can be improved. Its conveyance and gathering are expanded through a few procedures, for example, PC-supported programming, network frameworks, data sets, and data frameworks. Furthermore, the use of teaching aids helps stimulates interest and invigorates students thinking skills in this digital age and has been thought of and perceived as an impetus for the educating and growing experience. In many fields of computer science, artificial intelligence (AI) is a challenging and creative field. Using the computer-assisted instructional system (CAI) can provide an ideal instructional environment and easily stimulate students' enthusiasm and initiative to learn, thus significantly enhancing the instructional effect (Huang & Zhou, 2022). The increasing development of multimedia technology and integration with other leading technologies will undoubtedly encourage the further development of CAI. The Bayesian network approach is one of the artificial intelligence techniques often used in education. Bayesian Network is one of the applications of data mining that produces an interactive opportunity prediction model in the form of a description of the relationship between variables and provides information on probabilities.

Antecedent of Bayesian Network

A Bayesian network is an approach derived from the Bayesian theorem. The Bayesian network approach serves as an approach used in software development to determine the uncertainty measured using probabilities. Thomas Bayes put forward this theorem with the basic formula:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Bayesian Network is a Directed Acyclic Graph (DAG) equipped with a Conditional Probability distribution Table (CPT) for each node. Each node addresses a domain variable, and each arrow between nodes addresses a probabilistic reliance (Pearl, 1988). Bayesian networks can generally calculate the conditional probability by assigning values to other related nodes. Naive-Bayes Bayesian Network is a simple structure with nodes classified as parent nodes from several other nodes.

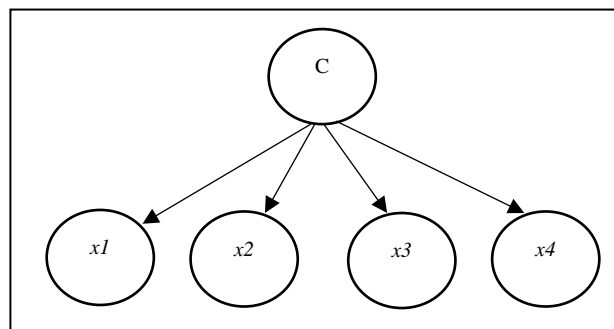


Figure 1. Bayesian network tree

Figure 1. Shows the structure of a Bayesian network classified using probability nodes. The calculation of the chance value at a node in the Bayesian Network structure is by the following formula:

$$P(X_1) = \sum_{j-i}^x P(X_1 - Y_1)$$

In conclusion, a Bayesian network is an approach that can represent uncertainty and reasoning in the field of artificial intelligence. Bayesian network is represented by an actual state and not a process of reasoning the probability of joint events between two events X_1 dan X_2 , where the probability of $P(X_1) > 0$ determined by the above formula.

Bayesian Network Structure

The Bayesian Network is a probabilistic graphical structure that illustrates causal relationships between interrelated variables. There are four things that Bayesian Network can offer as a method (Dita et al., 2018):

- i. Bayesian Networks can easily handle inaccuracies or problems with data.
- ii. The Bayesian Network allows one to learn about causal relationships. The learning process becomes essential when trying to understand the domain of a problem.
- iii. The Bayesian Network can facilitate a combination of domain and data knowledge.
- iv. Bayesian Network offers an efficient and principled approach to avoid overfitting the data.

Bayesian Network approach involves two steps which are creating the network structure and estimating the probability value of each node. One of the algorithms that can be used to form the network structure is to use the Maximum Spanning Tree algorithm. The Bayesian Network consists of two components, the Directed Acyclic Graph (DAG) and the Conditional Probability Table (CPT) for each attribute variable (Han & Kambe, 2001). DAG can be denoted by $G = (X, E)$, where X consists of variables called nodes and E is a pair of nodes connected by an arrow line. Nodes on the DAG represent random variables, while arrows indicate probabilistic dependence relationships between attributes. The arrow line only points in one direction and does not rotate back to its original node. The use of the Bayesian network in the development of software and applications in education is also seen to have a comprehensive impact. According to Dita et al., (2018) the Bayesian Network approach was used in developing an Intelligent Tutoring System (ITS) that can recommend materials appropriate to students' level of understanding. Meanwhile et al., (2015) used the Bayesian Network approach to identify the relationship between the factors of this study and analyse the data as it can represent variables in the form of nodes and relationships with directional lines. Constraint-based algorithms and score-based algorithms were used to generate networks in several categories to make comparisons and identify the factors that most influenced student learning in the subject of Additional Mathematics. Past studies have shown that this Bayesian Network approach is widely used for software development and educational applications. In line with that, this bibliometric analysis explores the literature on the Bayesian Network approach in education in the Scopus database. Furthermore, it aimed to answer the following research question:

1. How far has the Bayesian Network approach in education research progressed in the publication?
2. What is the scientific productivity pattern in the Bayesian Network approach in education field research?
3. What is the main area of the Bayesian Network approach in education research?
4. What is the future direction of the Bayesian Network approach in education?

It attempted to divulge publishing trends, patterns of scientific productivity, patterns of research conducted, and the primary area of the Bayesian Network approach in education.

Method

This study was conducted using the PRISMA approach (Preferred Reporting Items for Systematic Review and Meta-Analysis) and Bibliometric analysis to answer all research questions. The PRISMA approach is used to shortlist and select articles from Scopus databases based on the research topic. In addition, the PRISMA method has a specific procedure of how it is produced and a comprehensive scope that enfolds all appropriate materials and can replicate other researchers with the same approach to discussing a topic. Table 1 shows the inclusion and exclusion criteria for selecting and shortlisting the articles from the Scopus database.

Table 1. Inclusion and exclusion criteria

Inclusion	Exclusion
The article period of 2012 to 2021.	Duplicate article with same author and topic removed
The article focusses Bayesian network approach in education.	Bayesian network in other areas than education is excluded in this analysis.
English language papers only analysed in this study.	Other than English language excluded.

Data Collection

Data analysis begins with a keyword search by applying inclusion and exclusion criteria. Then, the keywords and search strings can be repeated to get the latest material and future research.

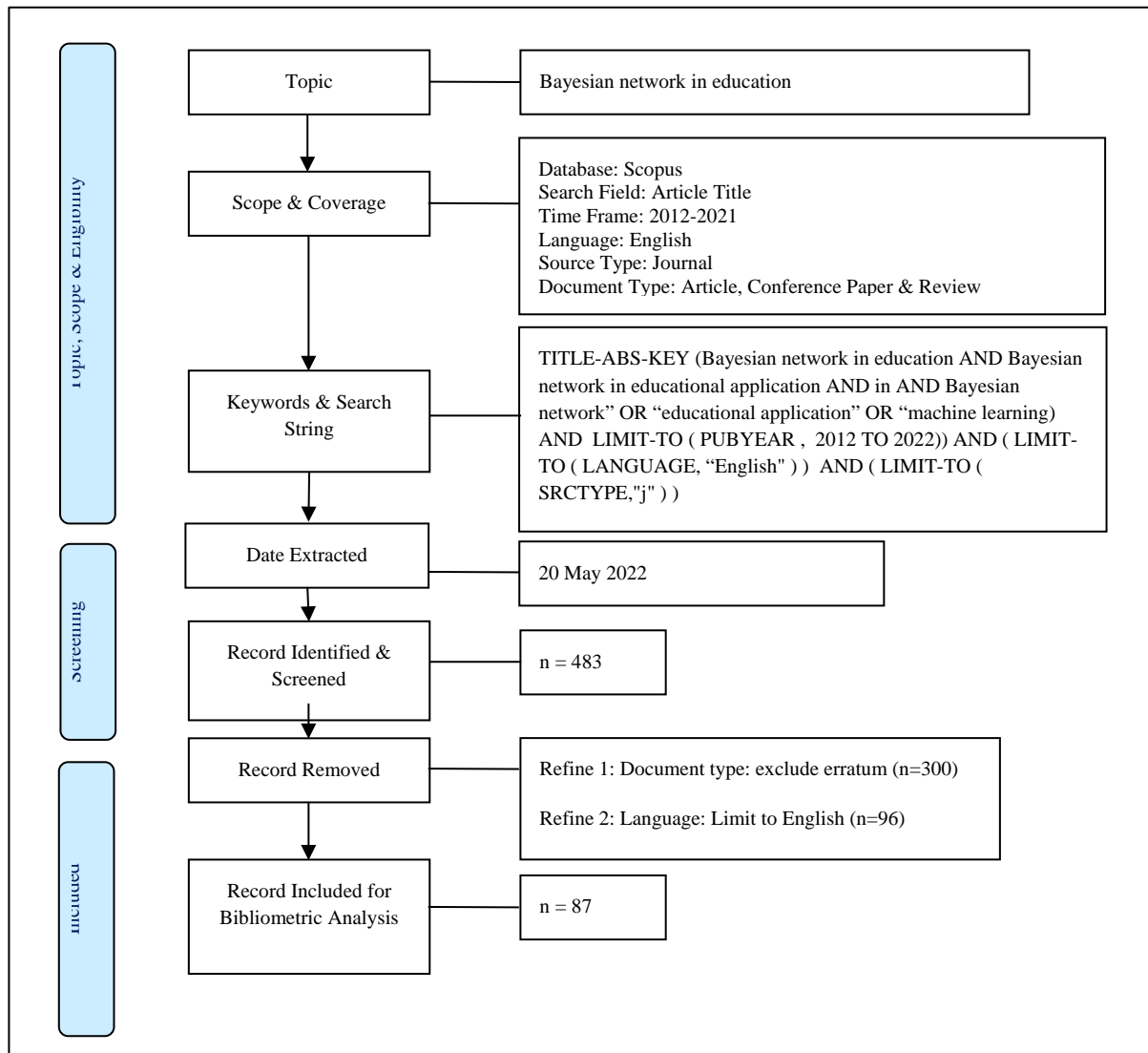


Figure 2. PRISMA flowchart of data inclusion and exclusion (Page et al., 2021)

This study included all the reports written in English from 2012 to 2022. Furthermore, excluding the erratum (n=300) to avoid twofold counting. Finally, 87 records are distinguished and downloaded for additional analysis. As in Table 2, three sets of keywords were used to select and shortlist the articles from the Scopus database. The primary keyword is the Bayesian network approach in education. While the secondary keyword is Bayesian network in educational application. The keywords used in this bibliometric analysis regarding Bayesian network and model are depicted in Table 2, in which the set of keywords are used with “AND” or “OR” operator.

Table 2. List of keywords

Keywords set	Keywords
Keywords_Set1	“Bayesian network in education” OR “Bayesian network”
Keywords_Set2	“Bayesian network in educational application”
Keywords_Set3	“Bayesian network” OR “educational application” OR “machine learning”

Data Analysis and Finding

In this study, performance and bibliometric analysis were blended to answer the research question. Based on Harzing (2007), the Bibliometrix R package and Perish software was used to conduct citation and publication analysis by running performance analysis. Meanwhile, VOSviewer were used to map the author's keyword. The usage of this software determined the significance of the study and involved prior Bayesian network clusters. On the other hand, the Total Link Strength (TLS) was used to decide the centrality of the study. The density of the scope is utilized to decide the research extension's internal strength or level of interaction within a network. Then again, the novelty was identified through the median value of the average publication. Average publication years greater or equal to the median value were considered a novel and the other way around.

Publication and Citation Trend

Table 3. Annual total citation

Year	N	TC	Mean TC per Art	Mean TC per Year	Citable Years
2012	4	224	56.00	5.60	10
2013	5	54	10.80	1.20	9
2014	4	4	1.00	0.13	8
2015	1	1	1.00	0.14	7
2016	5	71	14.20	2.37	6
2017	3	102	34.00	6.80	5
2018	5	8,3	1.67	0.42	4
2019	10	27	2.70	0.90	3
2020	20	20	1.00	0.50	2
2021	30	90	3.00	3.00	1
Total	87	601,3			

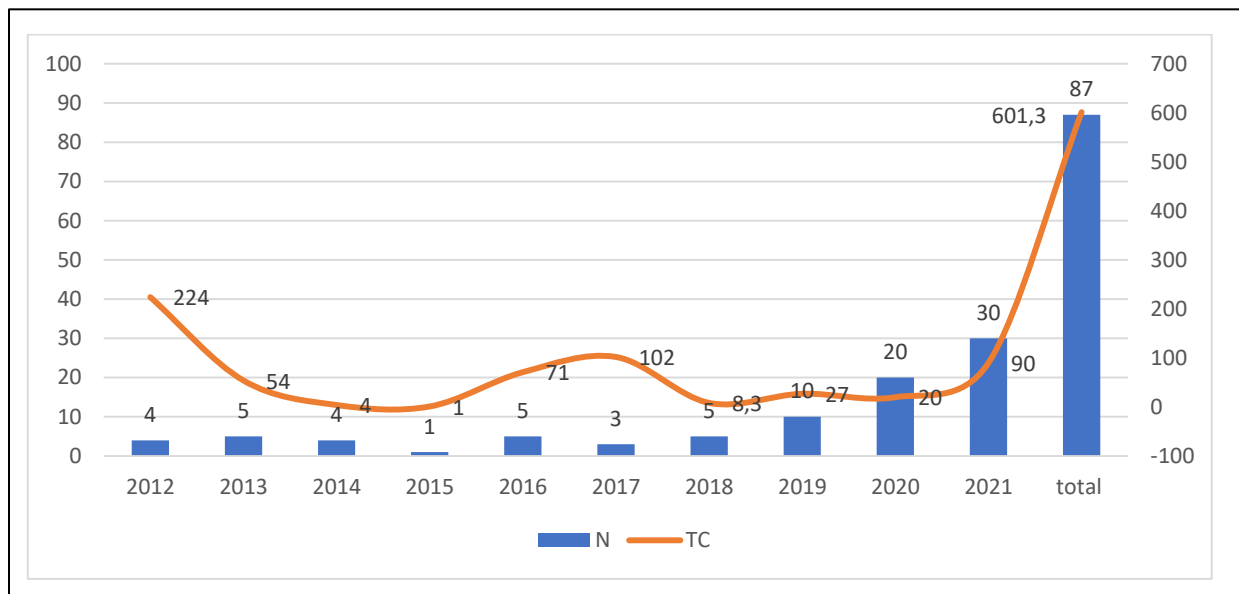


Figure 3. Total publication and citation on Bayesian network from 2012 to 2021

There were 87 publications on the Bayesian network in education research retrieved from the Scopus database for this study. The first publication for 2012, “Evaluation of simulation games for teaching engineering and manufacturing,” was included in the analysis (Hauge & Riedel, 2012). The number of publications related to the Bayesian network in education remained in the single digits from 2012 until 2018. However, the publications have been steadily increased since then (Table 3). The trend line shows that the number of publications increases polynomials ($R^2 = 601.33$), more significant than the linear increase.

Scientific Productivity Pattern

A total of 416 authors contributed to the publication of the Bayesian network in education research. Table 4 lists the number of publications based on countries has contributed. Based on the analysis, most authors have only published once on the topic of Bayesian networks in education. On the other hand, the United Kingdom became the country where it produced the most publication among the other 14 countries. Table 4 shows the United Kingdom published 18 publications, followed by United States with a complete publication of 16 publications, whereas Italy with seven publications.

Table 4. Top 10 countries contributed to publication on bayesian network in education

Country	Total Citations	Average Article Citations	Total Citation
United Kingdom	18	16	160
United States	16	14	179
Australia	5	7	115
Italy	7	6	58
Germany	6	5	78
China	3	1.5	50
South Korea	5	4	34
Spain	5	4	57
Netherland	5	3	20
Iran	5	3	0

Main Area of the Research on Bayesian Network in Education

The primary province of the Bayesian network in education research was distinguished utilizing keyword analysis. The analysis shows, only 49 of the 3249 keywords used by the authors outperformed the minimum occurrence level of 10 (Figure 4). Model, computer vision, structure, and machine learning are the main clusters that emerge from the map. The keyword overall strength was determined by generated the Enhanced Strategic Diagram (Figure 4). The link strength showed on centrality and meanwhile frequency showed by density. The average year of publication determined the novelty of publication. The high density shows publication on scope of child epidemiology, risk assessment of learning, besides interdisciplinary shows research on bayesian network in education related to students.

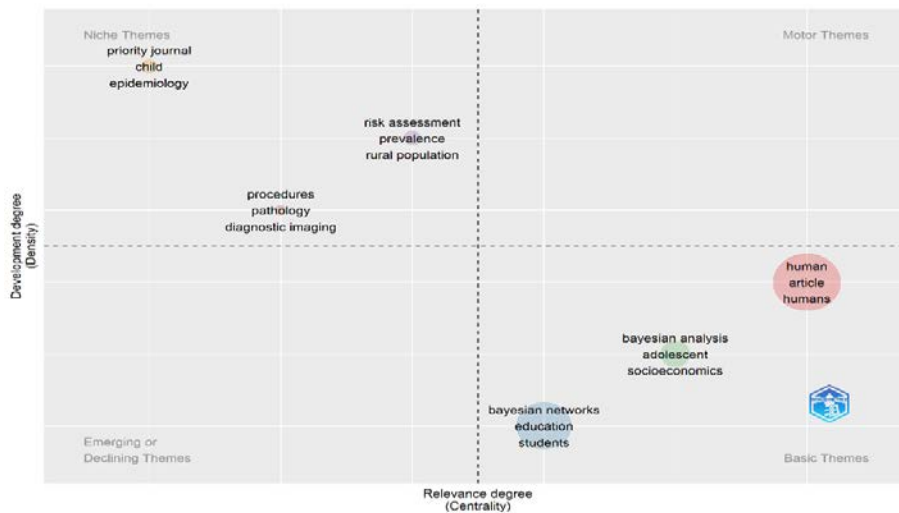


Figure 4. Enhanced Strategic Diagram (ESD)

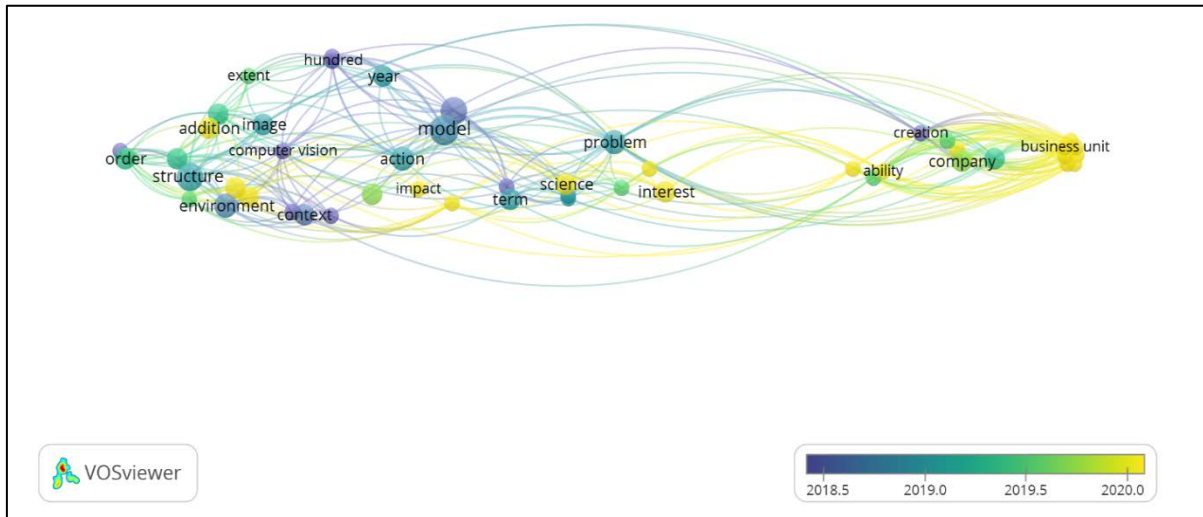


Figure 5. The co-occurrence of the author's keyword with minimum threshold of 10

Discussion and Future Direction

Bayesian networks were constructed in the current study of 'Probabilistic Arguments in Systems Intelligence' by Pearl (1988), which prompted the acceptance of probability and decision theory in artificial intelligence. The Bayesian Network officially allows for efficient representation and careful reasoning with definite science. Bayesian networks make it possible to be able to work from experience as well as combine the best artificial intelligence and neural networks. Bayesian networks are a family of probabilistic graph models. According to Devni et al., (2016), data mining in education has become widespread. Bayesian network, which is one of the data mining approaches, is also one of the main components inherited in data mining. Therefore, the use of the Bayesian network in education is increasing where the development of intelligence-based systems can facilitate educators' educational affairs. Furthermore et al., (2017) found that using the Bayesian network in the development of e-learning-based applications facilitates the ability to conclude the characteristics of individual users by receiving information data in the form of input from the user. Thus, the development of artificial intelligence applications in education can be realized using the Bayesian network approach. Moreover, the e-learning platform developed in the future can be expanded in all educational institutions using Bayesian Network. The Bayesian network approach was successfully implemented to generate individual student reports (Salwa et al., 2021). In this case, the method provides recommendations related to an e-learning session attended by students. Thus, the need to use the Bayesian network approach in education is increasing to guide students in making a decision related to learning. The finding can be girded by the increasing trend in implementing studies from 2018 to 2021 (Table 3). In particular, studies for 2019 and 2020 in Bayesian networks in education have increased. Thus, the future of the Bayesian field is envisioned to increase. This can be seen through the findings of previous studies that show the significance of the Bayesian network approach in the development of educational applications and software.

Conclusion

Overall, the findings of previous studies show that research trends in the field of Bayesian networks show an increased over the past five years. The use of the Bayesian network approach in the development of educational applications has also shown an increase. Bayesian networks are increasing in line with the expansion of artificial intelligence and data mining in education (Johnson et al., 2022). The increasing use of tablet and smartphone facilities among school students also contributes to developing applications using the Bayesian network approach and artificial intelligence that facilitate the decision-making process in terms of learning (Huang & Zhou, 2022). This study's findings describe the situation in using the Bayesian network in education, particularly aspects of trends, the number of publications, and the country published. It also discusses the future direction of the Bayesian network in education. However, the trend shows an increase in the use of Bayesian networks in education. There is still a need to expand research related to Bayesian networks in education. In particular, studies related to the importance and benefits of using Bayesian networks in studies should also be conducted. Studies on the importance and benefits can show productivity in using the Bayesian network approach in education.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSS journal belongs to the authors.

Acknowledgements

*This article was presented as an oral presentation at the International Conference on Social Science Studies (www.iconsos.net) conference held in Istanbul/Turkey on August 25-28, 2022

*We appreciate the support from FSKIK UPSI in preparing the journal of this article. The authors would also like to thank each individual involved in this research.

References

- Al-Qozani, H., & Aleryani, A. (2018). *The impact of iot on the higher education*. 2(2), 38–48.
- Azrin & Nurfaradilla. (2021). Kompetensi teknologi e-pengajaran guru bagi mendepani pendidikan pasca pandemik Covid-19. *Jurnal Penyelidikan Sains Sosial (JOSSR)*, 4 (11), 58 - 73.
- Belland, B. R., Walker, A. E., & Kim, N. J. (2017). A bayesian network meta-analysis to synthesize the influence of contexts of scaffolding use on cognitive outcomes in STEM education. *Review of Educational Research*, 87(6), 1042–1081. <https://doi.org/10.3102/0034654317723009>
- Cai, R., Wang, Q., Xu, J., & Zhou, L. (2020). Effectiveness of students' self-regulated learning during the Covid-19 pandemic. *Sci Insigt*, 34(1), 175-182. <https://doi.org/10.15354/si.20.ar011>
- Devni P. S., Dedi R., Danardono & Adhitya R. E., (2016). Pembentukan struktur bayesian network dari data. *Prosiding Konferensi Nasional Matematika XVIII*. 3-5 November 2016. universitas Riau, Pekanbaru.
- Dita, Wahyudin & Rasim. (2018). Pengembangan intelligent tutoring system menggunakan bayesian network. *Jurnal Teori Dan Aplikasi Ilmu Komputer*. 1(2). 61-71.
- Habibie. (2020). *Covid-19 corona virus response*. <http://mpoc.org.my/malaysian-palm-oil-industry/>
- Han & Kamber. (2001). Metarule-guided mining of multi-dimensional association rules using data cubes. *Int. Conf. Knowledge Discovery and Data Mining (KDD'01)*, 1(20). 207–210.
- Hasbi M. & Syarib M., (2017). Penerapan metode bayesian network dalam aplikasi e-learning berbasis web. jurnal sistem informasi, *Teknologi Informatika Dan Komputer*. 7(2), ISSN 2089-0265.
- Hauge, J.B., & Riedel, J.C. (2012). Evaluation of simulation games for teaching engineering and manufacturing *Procedia Computer Science*, 15, pp. 210-220.
- Huang, J. & Zhou, Q. (2022). Partitioned hybrid learning of Bayesian network structures. *Mach Learn*. 111, 1695–1738. <https://doi.org/10.1007/s10994-022-06145-4>
- Johnson, J.D., Smail, L., Corey, D., & Jarrah, A.M. (2022). Using Bayesian networks to provide educational implications: mobile learning and ethnomathematics to improve sustainability in mathematics education. *Sustainability*, 14, 5897. <https://doi.org/10.3390/su14105897>.
- Korganci, N., Miron, C., Dafinei, A., & Antohe, S. (2014). Comparison of generating concept maps and using concept maps on students' achievement. *eLearning and Software for Education*, 2, 287–293. <http://doi.org/10.12753/2066-026X-14-098>
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10, 325–337. <http://doi.org/10.1023/A:1008929526011>
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106, 901–918. <http://doi.org/10.1037/a0037123>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., & Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372:n71. <http://doi.org/10.1136/bmj>.
- Raja, R., & Nagasubramani, P. C. (2018). Impact of modern technology in education. *Journal of Applied and Advanced Research*, 3(S1), 33. <https://doi.org/10.21839/jaar.2018.v3is1.165>
- Teräs, M., Suoranta, J., Teräs, H., & Curcher, M. (2020). Post-covid-19 education and education technology 'solutionism': A seller's market. *Postdigital Science and Education*, 2(3), 863–878. <https://doi.org/10.1007/s42438-020-00164-x>
- UNESCO. (2019). *Artificial intelligence in education: challenges and opportunities for sustainable development*. <https://en.unesco.org/themes/education-policy>.

Author Information

Maran Chanthiran

Department of Computing, Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia
Contact e-mail: maranchanthiran@gmail.com

Abu Bakar Ibrahim

Department of Computing, Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia

Mohd Hishamuddin Abdul Rahman

Department of Computing, Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia

Punithavili Mariappan

Department of Arts, Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia

To cite this article:

Chanthiran, M., Ibrahim A. B., Abdul Rahman M. H., & Mariappan, P. (2022). Bayesian network approach in education: a bibliometric review using R-tool and future research directions. *The Eurasia Proceedings of Educational & Social Sciences (EPESS)*, 25, 17-25.