

Bayesian Network Approach in Educational Application Development: A Systematic Literature Review and Bibliometric Meta-Analysis

Maran Chanthiran¹, Abu Bakar Ibrahim¹, Mohd Hishamuddin Abdul Rahman¹,
Punithavili Mariappan^{1*}

¹ Department of Computing, Faculty of Arts, Computing and Creative Industry,
Universiti Pendidikan Sultan Idris. Tanjung Malim, Malaysia.

Article History

Received:
15.04.2022

Revised:
10.05.2022

Accepted:
20.05.2022

***Corresponding Author:**
Punithavili Mariappan
Email
mpunit.pm1111@gmail.com

This is an open access article,
licensed under: [CC-BY-SA](#)



Abstract: Technological developments have brought about a paradigm shift in the world of education. The education system must be more open and flexible, where students can experience these opportunities according to their skill level. 21st-century education and the application of the elements of Revolution 4.0 Industry in education realize that initiative. The Bayesian Network approach is becoming one of the essential tools in the development of educational applications. Therefore, the persistence of this systematic review is to identify peer-reviewed literature on the Bayesian network approach in education. Scopus and Web of Science, and IEEE citation databases are used in the data-gathering phase. PRISMA approach and keyword search were obtained and analyzed. This bibliographic data of articles published in the journals over ten years were extracted. VOS viewer was used to analyzing the data contained in all journals and articles. This systematic review shows that the development in education can absorb the changes that occur in technology. The findings from 1160 articles extracted show that using the Bayesian approach in the development of educational applications improves the quality of use, especially from the point of students. The level of predictive accuracy generated through the Bayesian network approach improves the quality of educational application development. However, the study's findings indicate that there is scope for research related to the application and use of this approach in the development of educational applications.

Keywords: Application Development, Artificial Intelligence, Bayesian Network, Bibliometric Analysis, Education.



1. Introduction

Mansor et al. [1] stated that integrating the latest technology with graphic elements can quickly improve students' understanding of a subject. The use of graphic elements in education requires minimal guidance from teachers. Additionally, Mansor et al. [1] emphasized that using the latest technological approaches such as artificial intelligence can improve student achievement in the subjects studied.

Leow & Neo [2] the use of multimedia elements in creating learning content makes the learning experience more meaningful. It becomes an essential learning component as it gives students an alternative method to have more choices when learning in a student-centred learning environment. Learning attitudes change when students realize that multimedia elements are more flexible in exploring and building new knowledge.

Finally, Saipullah & Hamzah [3] stated that developing and using multimedia concept applications can increase students' interest in learning Arabic. However, the readiness for use among teachers is still at a moderate level. The researcher also suggested providing training for teachers to improve their competence in handling multimedia teaching materials. Using technology models such as the Bayesian network helps designers and application developers develop functional instructional materials for students and teachers.

2. Antecedent of Bayesian Network

The Bayesian theory was adopted from its inventor, Thomas Bayes, around 1950, often found in statistical studies based on Bayesian theorems or rules. Bayesian theory is a theory of probability conditions that consider the probability of an event (hypothesis) depending on other events (evidence). In essence, the theorem says that future events can be predicted provided previous events have occurred [4].

Bayesian thinking provides a probabilistic approach to obtaining an inference or conclusion. The inference in a Bayesian Network is obtained from the relationship of each node in the Bayesian structure. Any change that occurs from a node will also affect the probability values of other nodes, which are directly or indirectly related to that node.

Bayesian Networks is a probability-based data modelling method that represents a set of variables and their conditional dependencies through a DAG (Directed Acyclic Graph). Each node formed in the graph has a Conditional Probability Table (CPT). In addition to CPT, the Bayesian Network model also has a Joint Probability Distribution (JPD), the probability of joint occurrence for all possible values contained in the variable. The formula used in Bayesian theory is as follows:

$$P(A|B) P(B) = P(A, B) = P(B|A) P(A) \text{ where } P(A|B) \text{ is probability of event occur from } A \text{ and } B.$$

Bayesian network is a classification with probability and statistical methods that predict future opportunities based on experience [5]. Bayesian networks perform probability calculations of all classes for all masks or features that make up a document. Furthermore, the Bayesian network selects the classification class based on the calculation of the probability of the class extract. The class with the most significant probability compared to the other classes will be selected as the class most eligible to be used [6]. In education, Bayesian networks are used in identifying changes in student behaviour in the classroom.

According to Afzal et al. [7], text visualization or even translating the text into graphic form became a relatively impossible thing before the existence of artificial intelligence components. Still, it can be made a reality with a combination of some of the latest technological techniques. This statement came true when Zou & Yue [8] developed a Bayesian Network model to plot the accident scene simply by entering some of the required information. Bayesian Network can predict traffic errors committed by road users by connecting nodes from text input.

According to Choon & Kumarenthiran [9], a Bayesian network is an approach to determining or making assumptions about student behaviour and attitude change. In line with that, Choon O. H and Kumarenthiran [9] were used to identify the factors that influence students in mastering the subject of Additional Mathematics in five National Type Secondary Schools in an urban area of Malaysia. The findings show that the probability is targeted on the factors that influence students attending Additional Mathematics classes.

Finally, Adiwisastra & Basjaruddin N. C [10] found that the Bayesian network can measure cognitive ability in basic Physics subjects. Researchers have developed an Intelligent Tutoring System (ITS) to measure the cognitive ability to master the subject of fundamental Physics. Overall, the trend shows that the Bayesian network is used to survey, measure, and identify changes in student behaviour towards learning a subject.

So, this systematic survey is to identify peer-reviewed literature on the field of text visualization in education through the Scopus, the Web of Science and IEEE database. The research work precisely by summarized the research objective as shown in Table 1.

Table 1. Research Question with its Significance

Research Objective	Significance of Research Question
- Analyse literature related to the use of Bayesian network in Education. - Identify trends in literature review of Bayesian network in educational application development. - Analyse the impact Bayesian network in education based on literature review.	- Contributed in identifying clusters that are frequently studied and be able to improve their related studies in the future. - Using the PRISMA approach and Bibliometric methods to identify research trends in the field. - Recommend the effectiveness of the use of Bayesian network in education, especially on student understanding and behaviours.

3. Research Methodology

Systematic literature review is desirable for research on this topic. As answer the research questions, PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) and Bibliometric analysis has been applied. PRISMA method applied to shortlist and selects the 1160 researches from Scopus, IEEE and the Web of Science databases with clear inclusion and exclusion criteria as below:

- The article period of 2010 to 2021
- Article focus on Bayesian network in Education
- Duplicate article with same author and topic removed

3.1. PRISMA Approach

A systematic review is the study of clearly constructed questions, using systematic and explicit methods to identify, select and critically evaluate relevant studies and collect and analyse data from selected studies [11]. The researcher able to show that the study is well-founded through a systematic study and identify gaps and directions that need to be addressed in future studies. This study is based on the PRISMA approach (Preferred Reporting Items for Systematic reviews and Meta Analyses). Based on this process and adaptation of PRISMA guidelines [12], this study is as follows:

- To answer the research question, (a) what is the latest cluster of Bayesian network studies in education, (b) what are the research trends in the topic study, (c) what is the impact of Bayesian network application in education and development of instructional materials.
- It is identifying and selecting significant studies using the PICOS framework.

Based on these guidelines, out of 1650 research papers identified (2010-2021), only 1160 research papers were used for the report after being screened and assessed for eligibility.

Table 2. PICOS Framework for Structuring Research and Analysis Questions

PICOS Component	Content
Population	Primary and Secondary Schools.
Intervention	Bayesian Network in the development of educational applications.
Comparison	Comparing Bayesian Network usage trends over a 10 -year period.
Outcome	Effectiveness of an intervention or study.
Study design	Review all empirical and analytical studies with a focus on the Bayesian Network approach in educational application development.

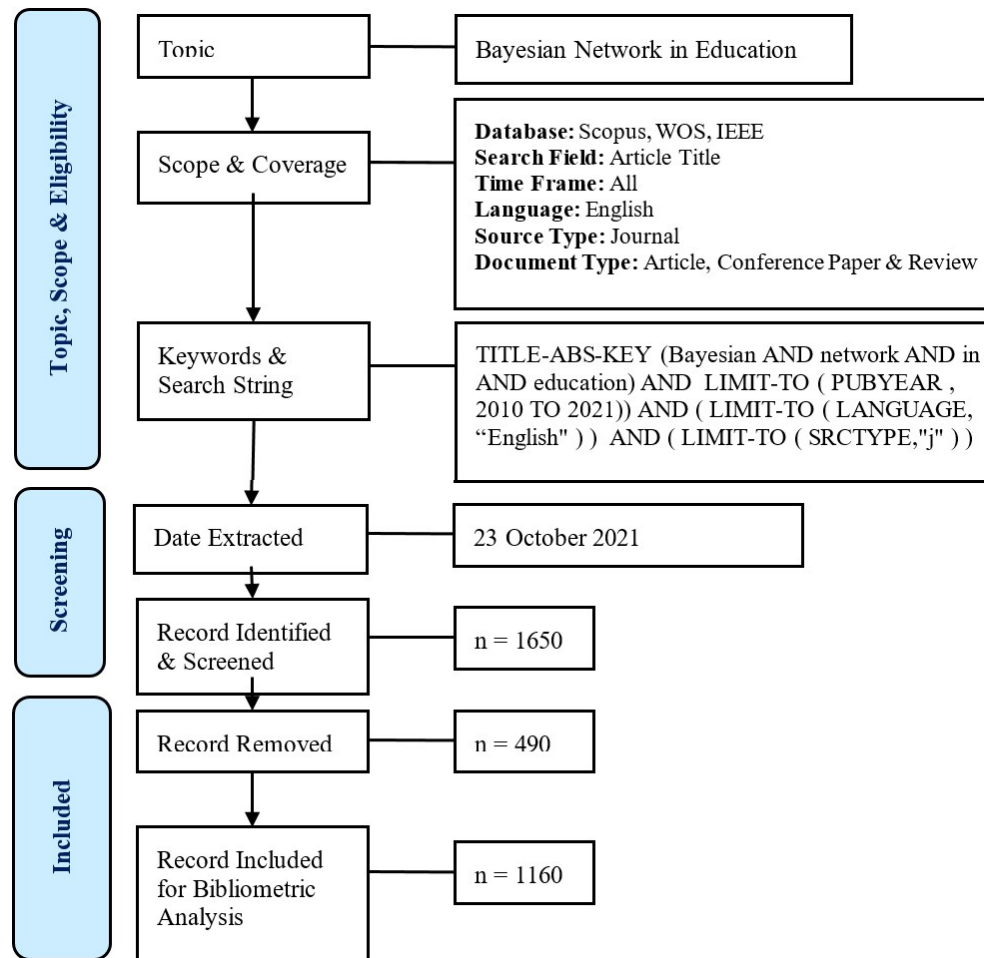


Figure 1. PRISMA Flowchart for Systematic Review of Bayesian Network in Education Bibliometric Analysis

3.2. Bibliometric Analysis

Bibliometric analysis was conducted to analyse the trend of the use of Bayesian network in education. Bibliometric analysis was carried out by filtering the article with some keywords as in the table. In the first phase, bibliometric analysis is carried out by analysing keywords. A total of five keywords were identified and analysed. In the second phase, an analysis of the publication was conducted based on a cluster from 1160 studies analysed.

3.2.1. Analysis of Keywords

A total of five sets of keywords were used in determining the study through the Scopus database, the Web of Science and IEEE. The primary keyword used is Bayesian network and education. While the secondary keywords used are model and development.

The keywords used in this bibliometric analysis regarding Bayesian network and model are depicted in Table 3, in which the set of keywords are used with “AND” or “OR” operator.

Table 3. List of all Keywords

Keywords Set	Keywords	Occurrences	Relevance
Keywords_Set1	“Bayesian network” OR “Bayesian” OR “applicability”	40	2.19
Keywords_Set2	“model” OR “computer graphic” OR “computer” OR “graphic”	45	1.26
Keywords_Set3	“teaching” OR “software” OR “development” OR “animation” OR “internet”	80	2.10
Keywords_Set4	“predictor” OR “education” OR “challenge”	54	0.70
Keywords_Set5	“education” OR “teaching tool” OR “trend”	79	0.98

3. Finding and Discussion

To analysis, the research questions formulated in the previous section, the study of this paper appropriated the accompanying parts of scholarly works. A total of 1160 articles were analysed in this bibliometric and SLR based on Scopus, IEEE, and WOS databases. VOS Viewer was utilized to investigate and answer the objectives of the study. The study was conducted to distinguish the frequency of publication of studies in the Bayesian network in Education. Among the aspects identified are publications issued from 2010 to 2021. Next, subject area and publication trends are also analysed in this literature review.

3.1. Subject Are of Bayesian Network in Education

Table 4 below shows the subject area that many studies were conducted during the period 2010 to 2021 in the scope of Bayesian network study. Findings show that most studies were conducted in computer science with 66.90%, which is 776 studies out of 1160. Meanwhile, the use of the Bayesian network in Education, especially the subject of Engineering, is also widely studied, with several studies of 304 (26.21%). On the other hand, veterinary subjects, Immunology, and Microbiology are subjects that are rarely studied.

Table 4. Subject Area

Subject Area	Total Publications (TP)	Percentage (%)
Agricultural and Biological Sciences	18	1.55%
Arts and Humanities	16	1.38%
Biochemistry, Genetics and Molecular Biology	23	1.98%
Business, Management and Accounting	34	2.93%
Chemical Engineering	8	0.69%
Chemistry	8	0.69%
Computer Science	776	66.90%
Decision Sciences	56	4.83%
Dentistry	2	0.17%
Earth and Planetary Sciences	10	0.86%
Economics, Econometrics and Finance	16	1.38%
Energy	21	1.81%
Engineering	304	26.21%
Environmental Science	40	3.45%
Health Professions	14	1.21%
Immunology and Microbiology	1	0.09%
Materials Science	25	2.16%

Mathematics	262	22.59%
Medicine	126	10.86%
Multidisciplinary	12	1.03%
Neuroscience	21	1.81%
Nursing	5	0.43%
Pharmacology, Toxicology and Pharmaceutics	8	0.69%
Physics and Astronomy	35	3.02%
Psychology	17	1.47%
Social Sciences	204	17.59%
Veterinary	1	0.09%

Table 5 shows the 30 most keyword analyses in 1160 edited study articles. Findings show that the keyword "Bayesian networks" is widely used in 832 publications which are 71.98%, while "education" is used in 497 publications (42.84%). In addition, "Inference engines" and "Learning systems" are also widely used as keywords.

Table 5. Keyword Analysis

Author Keywords	Total Publications (TP)	Percentage (%)
Bayesian Networks	835	71.98%
Education	497	42.84%
Inference Engines	240	20.69%
Students	234	20.17%
Learning Systems	233	20.09%
Education Computing	199	17.16%
Artificial Intelligence	173	14.91%
Bayesian Network	171	14.74%
Learning Algorithms	162	13.97%
Intelligent Networks	140	12.07%
Distributed Parameter Networks	138	11.90%
Human	128	11.03%
Data Mining	120	10.34%
Article	118	10.17%
Computer Aided Instruction	111	9.57%
Engineering Education	104	8.97%
Humans	104	8.97%
Bayes Theorem	98	8.45%
E-learning	98	8.45%
Teaching	92	7.93%
Forecasting	84	7.24%
Intelligent Tutoring System	82	7.07%
Bayesian	80	6.90%
Machine Learning	79	6.81%
Neural Networks	75	6.47%
Speech Analysis	74	6.38%
Algorithms	67	5.78%
Bayesian Networks	835	71.98%
Education	497	42.84%
Inference Engines	240	20.69%

Next, an analysis is conducted to see the publication by trend or landscape of the study is conducted Table 6 shows the year of publication in the field of study where 2010 found a total of 212 articles published with indexed paper h of 13 and g indexed paper is 31 For the next year which is 2011 shows a total of 157 published with g and h indexed papers respectively 13 and 25. On the other hand, the year 2021 shows a total of 67 studies conducted in the Bayesian network in Education and only two papers h and g indexed. Overall, it can be concluded that studies related to the Bayesian network in Education have been inconsistent and declining over the past five years.

Table 6. Year of Publication

Year	TP	NCP	TC	C/P	C/CP	h	g
2021	67	15	20	0.30	1.33	2	2
2020	93	47	136	1.46	2.89	6	7
2019	84	54	372	4.43	6.89	10	17
2018	68	51	559	8.22	10.96	12	21
2017	110	88	883	8.03	10.03	17	25
2016	58	45	477	8.22	10.60	12	20
2015	82	40	888	10.83	22.20	14	29
2014	78	34	601	7.71	17.68	12	24
2013	76	32	393	5.17	12.28	12	19
2012	75	27	592	7.89	21.93	9	24
2011	157	34	688	4.38	20.24	13	25
2010	212	35	1012	4.77	28.91	13	31
Total	1160						

Notes: TP=total number of publications; NCP=number of cited publications; TC=total citations; C/P=average citations per publication; C/CP=average citations per cited publication; h = h -index; and g = g -index.

3.2. Visualisation Map

The following is a visualization map showing in detail the author keyword term co-occurrence network based on title and abstract fields (Binary Counting).

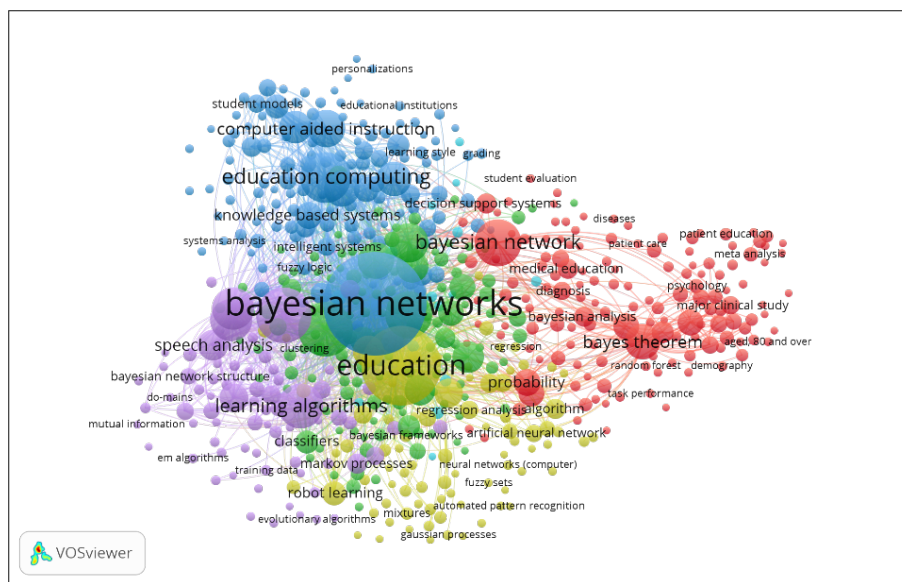


Figure 2. Network Visualisation Map of the Author Keywords

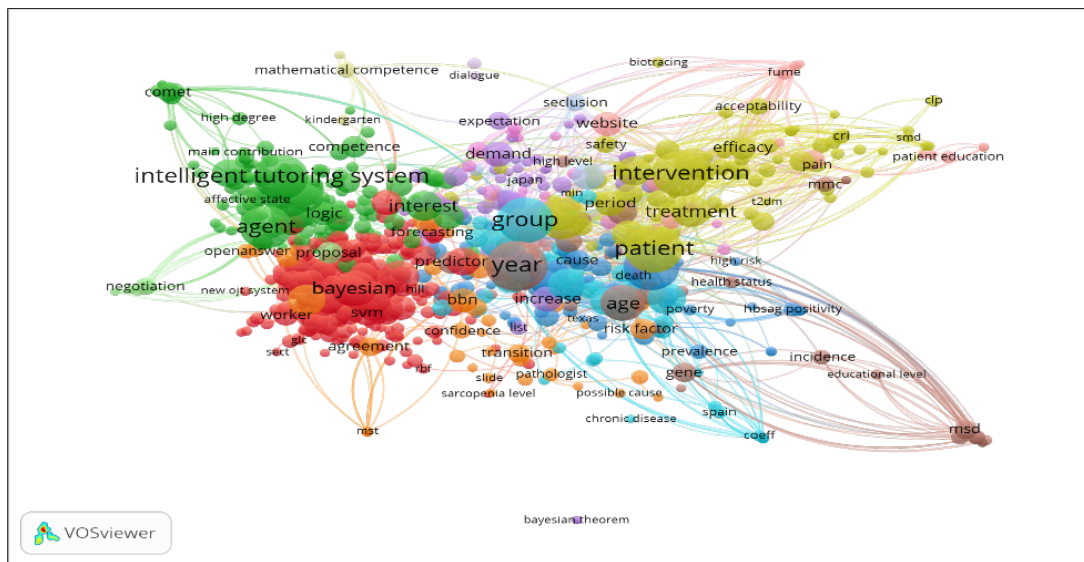


Figure 3. VOS Viewer Visualization of a Term Co-Occurrence Network Based On Title and Abstract Fields (Binary Counting)

4. Discussion

This systematic literature review using PRISMA and bibliometric methods has answered three research questions developed. In total, 1160 studies published in Scopus, IEEE and Web of Science were analysed. A total of five sets of keywords were published in screening the content of the study. Bibliometric analysis has provided an overview of the trends and effects of Bayesian network in education on educational application development and students behaviour. Overall, many studies are conducted in Education; however, studies in the last five years have decreased compared to the previous year. However, Rodriguez-Barrios et al. [13] stated that this Bayesian network is an effective model for analysing student behaviour changes, especially in analysing attitudes towards reading skills. The study shows that The Bayesian network model subtracts for discovering the language and communication level of performance based on learning style, learning pace, speed, and reading comprehension, obtaining an 85.36% accuracy. Next, Bayesian network model facilitates assessing learning behaviour among students [14] [15]. The study's findings indicate that the process of domain-related erudition modelling is the Bayesian networking process of the directions that students have learned. Secondly, the time and space complexity of the algorithm is reduced by interpreting the network structure and optimizing the order of node deletion in the triangulation process. Overall, many studies are conducted related to the use of Bayesian network usage in educational application development. Usage trends show that most studies are related to the development of applications and software and Bayesian network models in Education. Also, the trend shows studies related to the development of language and students' behavioural analysis applications in Education.

5. Conclusion

Studies on the use of Bayesian network in educational application development through systematic literature review and bibliometric meta-analysis show some recommendations in the future:

1. Studies related to the effectiveness of the use of Bayesian network in education show shortcomings in pass five years. Therefore, there is a need to increase research related to the effectiveness of the use of Bayesian network in education especially student behaviour analysis.
2. The use of the Bayesian network in the development of educational applications provides an accurate view, and it helps teachers identify students accurately in behaviour change.
3. Through, this Systematic literature review can be used as a guide in publishing new articles on the Bayesian network to develop educational applications.

This analysis evaluates the Bayesian network and its usage in educational application development. This review assesses the effectiveness and trend of using the Bayesian network. A total of 1160 studies cited from Scopus, Web of Science IEEE were systematically and bibliometric analysed. The findings conclude that the use of Bayesian models in developing instructional applications and software assists teachers in identifying and measuring changes in student behaviour in particular. In addition, the use of the Bayesian network also opens space for teachers to know students more accurately in terms of factors that contribute to increased involvement in teaching and learning activities.

References

- [1] N. R. Mansor, R. Zakaria, R. Abd. Rashid, R. Mohd Arifin, B. Hairiel Abd Rahim, R. Zakaria, and M. T. Mohd, "A Review Survey on the Use Computer Animation," in *Education IOP Conference Series: Materials Science and Engineering*, 2020. doi: 10.1088/1757-899X/917/1/012021.
- [2] F. T. Leow and M. Neo, "Interactive Multimedia Learning: Innovating Classroom Education in A Malaysian University," *Turkish Online Journal of Educational Technology*, vol. 13, no. 2, pp. 99–110, 2014.
- [3] N. A. F. Saipullah, and M. I. Hamzah, "Tahap Penggunaan Multimedia Dalam Pengajaran Bidang Al-Quran: Satu Kajian Rintis Terhadap Guru Pendidikan Islam Di Sarawak. Persidangan Antarabangsa Sains Sosial Dan Kemanusiaan Ke-5 (PASAK5 2020), 5(November)," *Journal of Quran Sunnah Education and Special Needs*, vol. 5, pp. 223–233, 2020.
- [4] Sugandi and Yahdin, "Aplikasi Pengambilan Keputusan Pada Perencanaan Produksi Berdasarkan Teorema Bayes," *Media Informatika*, vol. 1, no. 2, pp. 25-38, 2008.
- [5] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *Jurnal Informatika*, vol. 8, no. 1, pp. 884-898, 2014.
- [6] Manning, Raghavan and Schutze, *Introduction to Information Retrieval*. Cambridge University Press: USA, 2009.
- [7] S. Afzal, R. Maciejewski, Y. Jang, N. Elmqvist and D. S. Ebert, "Spatial text visualization using automatic typographic maps." in *Transactions on Visualization & Computer Graphics*, vol. 18, no. 12, pp. 2556-2564, 2012, IEEE.
- [8] X. Zou, and W. L. Yue, "A Bayesian Network Approach To Causation Analysis of Road Accidents Using Netica," *Journal of Advanced Transportation*, 2017.
- [9] O. H. Choon and Kumarethiran, "A Bayesian Network Approach To Identify Factors Affecting Learning of Additional Mathematics," *Jurnal Pendidikan Malaysia*, vol. 40, no. 2, pp. 185-192, 2015.
- [10] M. F. Adiwisatra and N. C. Basjaruddin, "Intelligent Tutoring System Untuk Mengukur Kemampuan Kognitif Dalam Fisika Dasar Berbasis Metode Bayesian Network," *Indonesian Journal on Computer and Information Technology*, vol. 2, no. 2, pp. 40-47, 2017.
- [11] D. Moher, L. Shamseer and M. Clarke, "Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) Statement," *Systematic reviews*, vol. 4, no. 1, pp. 1-17, 2015.
- [12] A. Liberati, D. G. Altman, and J. Tetzlaff, "The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration," *Annals of Internal Medicine*, vol. 15, no. 1, pp. 65-94, 2009.
- [13] E. U. Rodriguez-Barrios, R. A. I. Melendez-Armenta, S. G. Garcia-Aburto, M. Lavoignet-Ruiz, L. C. Sandoval-Herazo, A. Molina-Navarro, and L. A. Morales-Rosales, "Bayesian Approach to Analyze Reading Comprehension: A Case Study in Elementary School Children in Mexico," *Sustainability*, vol. 13, no. 8, pp. 4285, 2021.
- [14] S. Li, "Bayesian Network Algorithms Used in the Assessment of Learners' Learning Behaviour," *International Journal of Continuing Engineering Education and Life Long Learning*, vol. 3, no. 3, pp. 12-22, 2021.
- [15] S. M. Kanti, S. Dutta, and J. I. Laskar, "A hierarchical bayesian network model for flood resilience quantification of housing infrastructure systems," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, vol. 7, no. 1, 2021.