



Poulami Pal
Research Scholar
Dept. of Library and
Information Science,
University of Kalyani
palpoulami113@gmail.com

Defining Interdisciplinarity through Topic Modelling Technique: An AI-Driven Approach

Poulami Pal
Sibsankar Jana



Dr. Sibsankar Jana
HOD & Associate Professor
Dept. of Library and
Information Science,
University of Kalyani
sibsankarjana04@gmail.com

Abstract:

Interdisciplinary studies or research is a method of integrating knowledge and perspective from multiple disciplines to address complex problems. Interdisciplinarity consists of collaboration and cooperation from different fields to assemble together diverse perspectives and approaches in order to find innovative solutions. In this paper, we have measured interdisciplinarity using topic modelling approach. Topic modelling is an unsupervised natural language processing (NLP) technique for creating structured data from a collection of unstructured textual corpus without predefined training data. This text mining method have discovered hidden semantic patterns demonstrated by a text corpus to produce cluster of words, called "topics".

In this article, we have collected data from The Lens database (www.lens.org) and then applied topic modeling approach using the tool Coconut Libtool (www.coconut-libtool.com/). The final output is shown by network graphs, sunburst visualization, chord diagram etc. The main objective of the paper is to define the perspective of interdisciplinarity through topic model techniques.

Keywords: *Interdisciplinarity, topic modeling, Natural Language Processing, Text mining, Text analytical tool, Visualization.*

1. Introduction:

Interdisciplinary studies or research is a method of integrating knowledge and perspective from multiple disciplines to address complex problems. Interdisciplinarity consists of collaboration and cooperation from different fields to assemble together diverse perspectives and approaches in order to find innovative solutions. Text-based measures of interdisciplinarity are used to estimate the extent to which a document or a body of work integrates concepts, theories, and methodologies from multiple disciplines.

In this paper, topic modeling techniques are used to define text-based interdisciplinarity. Topic modeling is an unsupervised natural language processing (NLP) technique for creating structured data from a collection of unstructured textual corpus without predefined training data. This text mining method have discovered hidden semantic patterns demonstrated by a text corpus to produce cluster of words, called "topics". The basic idea of topic modeling is that each document is regarded as mixture of topics and each word of a document has a certain probability of



belonging to a specific topic. The topic-associated words can help to organize and uncover the insight into large amounts of unstructured textual data (Verma & Yuvaraj, 2023). Two popular topic model techniques are *Latent Dirichlet Allocation (LDA)* and *Latent Semantic Analysis (LSA)*. LDA is a generative probabilistic model that allocates words to topics and topics to documents. LSA applies a “term-document matrix” and “bag of words” to detect topics. There are many other techniques such as *Lda2vec*, Probabilistic Latent Semantic Analysis (pLSA) etc. In this paper, topic model techniques are done by using a tool, *Coconut libtool* (www.coconut-libtool.com), which is the first open-source web application that uses state-of-the-art natural language processing (NLP) technologies, integrating AI, LLM, NLP and machine learning algorithm along with purposeful designing model in the backend (Santosa et al., 2024).

2. Literature Review:

In this paper, the concept of interdisciplinarity of research or studies are discussed by using two web-apps- the Coconut Libtool and the Flourish, though there are many ways to define and measured interdisciplinarity of research. Evans (2016) discussed about text-based measurement of interdisciplinarity by comparing the text of a scholar’s publications with text from multiple disciplines (Evans, 2016). Kherwa & Bansal (2018) defined topic modeling technique as a statistical technique for revealing the underlying semantic structure in large collection of documents (Kherwa & Bansal, 2018). Saha (2021) proposed topic modeling as an efficient method for an exploratory literature review in management research using the Latent Dirichlet Allocation (LDA) method(Saha, 2021). Paul & Girju (2012) addressed the problem of scientific research analysis using the topic model LDA based on topic and language(Paul & Girju, 2012). Singh et al. (2023) defined the evolving relationship of entrepreneurship, technology, and innovation from topic modeling perspective (Singh et al., 2023). Laureate et al. (2023) gave a systematic review of the use of topic models for short text social media analysis which yield meaningful insights.(Laureate et al., 2023). Kim et al. (2024) proposed the application of topic modeling techniques to identify new emerging science and discussed about potential applications to identify emergence through the merging of global interdisciplinary domains(Kim et al., 2024). Yu & Xiang (2023) discovered topics and trends in the field of artificial intelligence by utilizing LDA topic modeling technique(Yu & Xiang, 2023).

3. Research Gap:

On the basis of existing literatures, it is clear that the application of topic modeling approach has been applying previously in various context and with many purposes. Though a large amount of work has been done on different aspects of discovering and measuring interdisciplinarity, a research gap is being noticed that usefulness of such user-friendly tools to find and discover the



interdisciplinarity. Also, these newly developed applications of utilizing of these tools are not done so much.

4. Immersion of the Problem:

- To discover common themes and concepts by utilizing topic model approach,
- To assess the diversity of topics within a corpus or text-based data,
- To monitor how interdisciplinary themes evolve over time into the dynamic nature of research fields.

5. Rationale of the Study:

Topic modeling technique helps uncover themes and insights, highlighting how different disciplines might intersect, indicates interdisciplinary connections. It also aids in conducting comprehensive literature reviews by clustering research papers into relevant topics. Topic modeling can track how interdisciplinarity evolves over time. It helps in evaluating the impact of interdisciplinary research by analyzing the prevalence and influence of interdisciplinary topics in the literature, discovering new knowledge at the intersection of different disciplines, which can lead to innovative approaches and solutions.

6. Methodology:

In this paper, topic modeling approach is used to discover interdisciplinarity in research. Here, a dataset of 1000 research articles are drawn from the Lens (www.lens.org) database as sample to find the interdisciplinarity and data analyses are done by using the user-friendly, open-source, web-based tools Coconut Libtool and Flourish. The Coconut libtool is a freely available data mining and text analysis tool. It offers data visualization and analysis of textual unstructured data. Flourish is an influencing data visualization and storytelling tool which enables the creation of interactive maps, charts directly from csv dataset. At first the file of the downloaded dataset is imported to the Coconut libtool and checked it out for the which types of analyses are supported. Then topic modeling, bidirected network, sunburst visualization, burst detection, scatter-text are deriving one-by-one using the tool. After that, the list of most 'bursty' or occurred topics are downloaded and demonstrated by chord diagram, which is created by using Flourish tool. This tool allows the csv file containing the list of words. At first, from the browser and then signed in to the web-app by username and password. Then the csv file is imported and chord diagram are generated automatically. Then it can be downloaded or saved.

7. Study Conducted:

This study conducted for interpreting or discovering interdisciplinarity using topic modeling method. So, dataset is collected from secondary source (lens.org) and then analyzed and

visualized by NLP technique and machine learning algorithms. Using AI-based tools the dataset is analysed in back end and visualized by diagrams, charts, maps, networks etc.

8. Data Collection:

The dataset is collected from The Lens database (www.lens.org) against the string “artificial AND intelligence” from scholarly works field and top 1000 documents are selected according to the relevance of the keyword. The excel file containing the 1000 documents are retrieved from the database.

9. Data Analysis:

The analysis of data are done by using two web-based tools- Coconut Libtool (www.coconutlibtool.com/) and Flourish (www.app.flourish.studio/). The Coconut Libtool uses natural language process (NLP) to deploy Topic Modeling, Bidirected Network, Sunburst Visualization, Burst Detection and Scatter-text. Using the web-application Flourish, Chord Diagram has been drawn to exhibit Interdisciplinarity.

- i) **Topic modeling:** In topic modeling section, there are three types of algorithms are used viz. PyLDA, Biterm and BERTopic. In each algorithm 10 numbers of topic were selected.

Using **PyLDA algorithm**, obtained Coherence score is 0.37836954305397386. Coherence score is a metric that measures the semantic similarity between high-scoring words in topic. A higher coherence score represents a better quality of model.

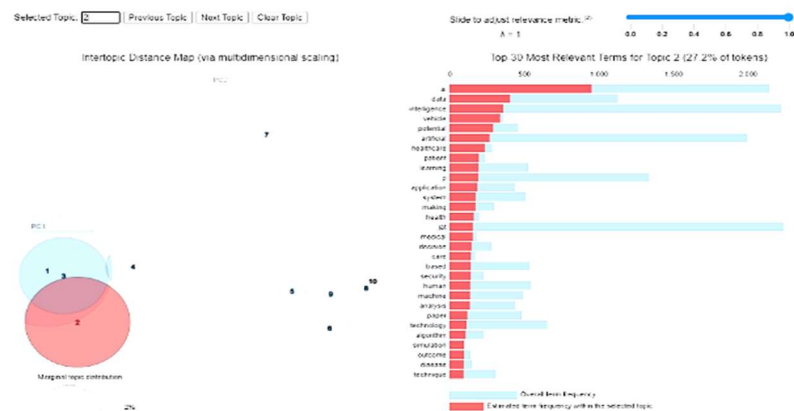


Figure No-1: Intertopic distance map (PyLDA algorithm)

By using **Biterm method**, derived Perplexity Score is 1224.7002752829987. Perplexity score is a metric that measures how well a model predicts new or unseen data. Perplexity score means the accuracy and confident of the model in its predictions.

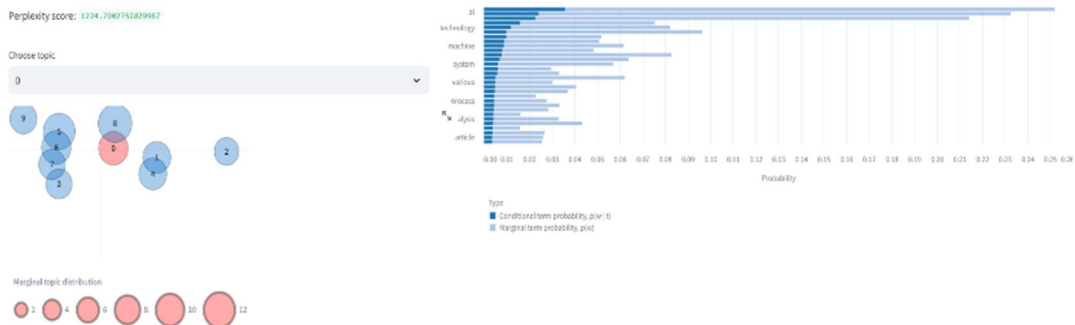


Figure No-2: Intertopic distance map (Biterm method)

BERTopic model visualizes topic, terms, documents, document hierarchy, topic similarity by intertopic distance map, topic word scores, documents and topics, similarity matrix and hierarchical clustering respectively.

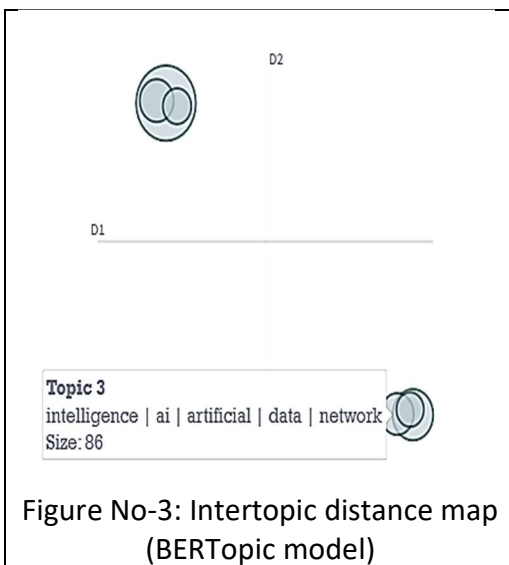


Figure No-3: Intertopic distance map (BERTopic model)

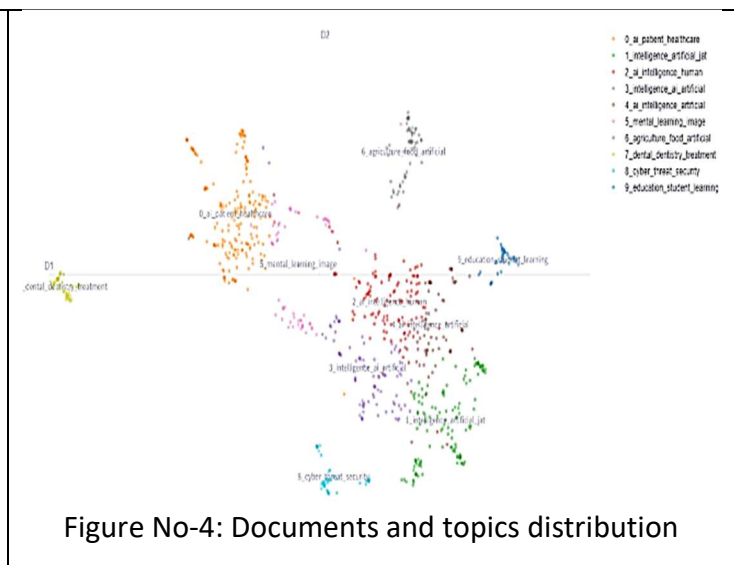


Figure No-4: Documents and topics distribution

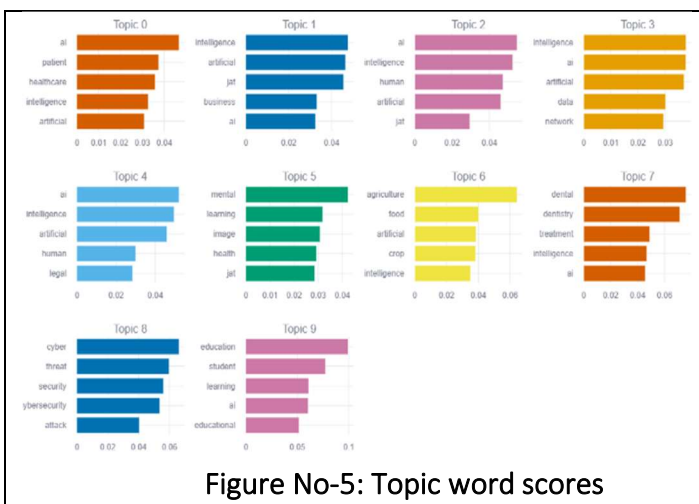


Figure No-5: Topic word scores

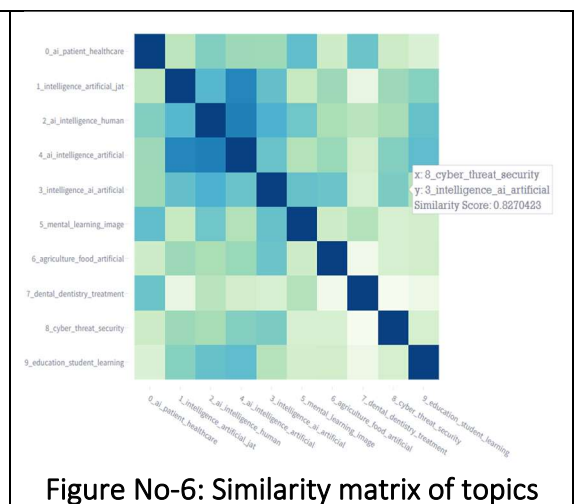
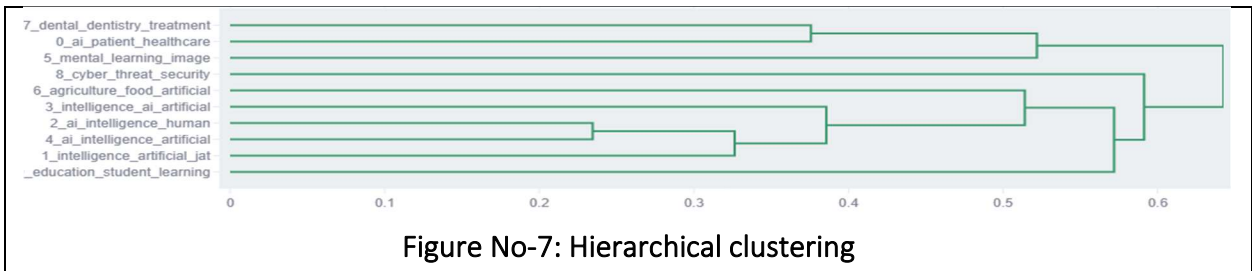
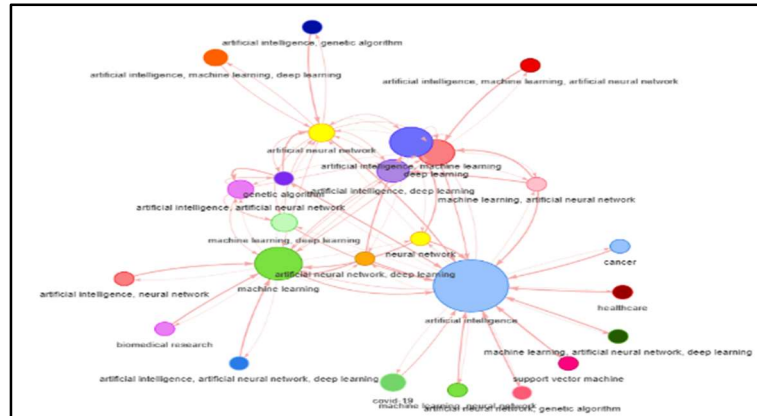


Figure No-6: Similarity matrix of topics

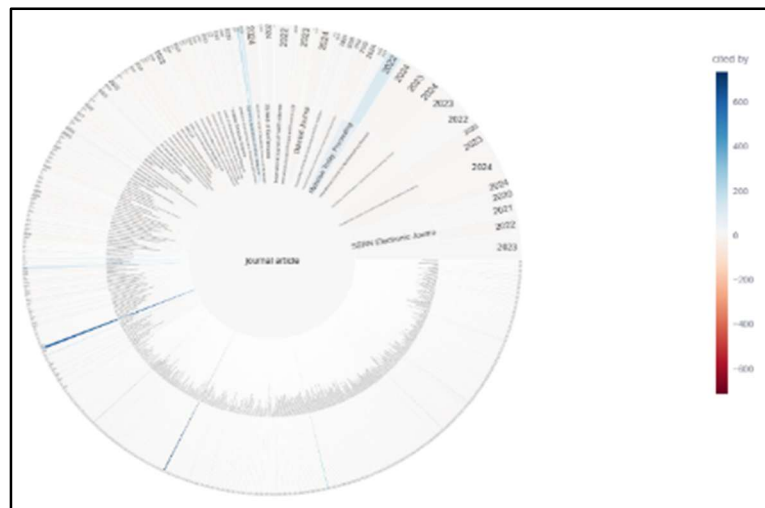


ii. Bidirected Network

Bidirected network represents data transmission in both directions between nodes. Here, the nodes are the topics extracted from documents.



iii. Sunburst Visualization: This displays hierarchical data. Each level of hierarchy is represented by one ring or circle and the innermost circle is the top of the hierarchy. Here, outer ring is displaying publication year, middle ring is representing the source title and inner ring represents document type. The following diagram depicts journal article as document type.





iv. Burst Detection: This identifies time periods in which a target topic is uncharacteristically frequent or 'bursty'. This indicated by a contiguous sequence of words whose occurrence have unusually higher. Here, it can be shown that the term "ai" is mostly occurred in 2023 and others most bursty words are "artificial" and "intelligence".

machine	20	56	147	147	147
research	21	43	108	160	133
learning	28	71	111	165	157
study	33	29	125	181	175
human	27	58	138	185	154
datum	37	49	122	203	171
technology	38	98	142	236	234
artificial	121	218	479	636	538
intelligence	127	225	484	726	586
ai	127	178	356	861	794
	2019	2020	2021	2022	2023

Figure No-10: Burst score of most used topics

v. Scatter-text: It is mainly used to explore data and find insights and add contextual elements like averages of frequency etc. The top-right corner indicates the most frequently occurring topics in documents and down-left corner represents less occurred topics in set of documents

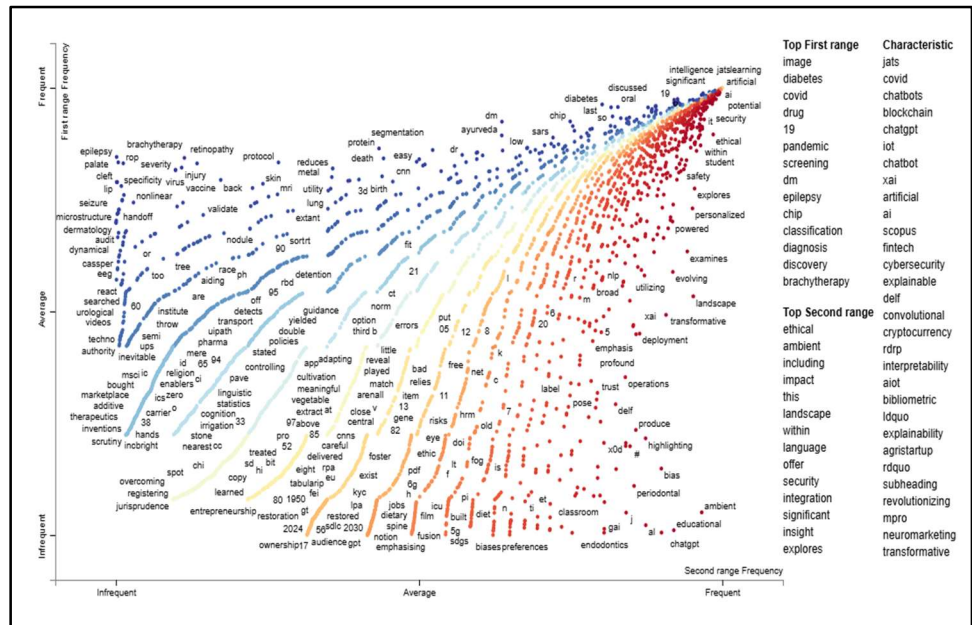


Figure No-11: Scatter diagram of keywords

Chord Diagram:

By using the web-app Flourish tool, the chord diagrams are developed. Chord diagram is a graphical method of displaying inter-relationships between data in a matrix. The arcs represent the importance of flow and size of arc depends on value. Here, the top keywords are arranged in recent two years used in documents of various subject and distribution of the topics are demonstrated by chord diagram.

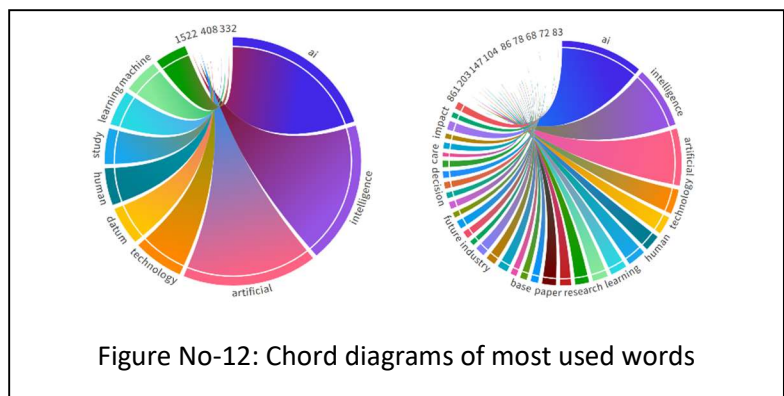


Figure No-12: Chord diagrams of most used words



10. Interpretation of Findings revealed from the Data Analysis:

Using the AI-based tools, the data analysis is conducted and represented by visualization and storytelling approach like chord diagram, bidirectional network, sunburst visualization, scatter text diagram etc. Therefore, topic modeling is a highly valuable tool for identifying overlapping themes, enhancing knowledge integration, facilitating collaboration and trend analysis.

11. Conclusion:

In conclusion, utilizing as an ai-driven approach, topic modeling can be better understandable and significantly enhanced the measurement of interdisciplinarity across diverse field of research or studies. Academic journals can use topic modeling technique to arrange special issues on emerging interdisciplinary topics. It can be concluded that topic modeling is such a powerful tool that helps to find interdisciplinarity, enable deeper insight and foster collaboration.

12. Delimitation of the Study:

Here, data mining and text analysis has been done by using data visualization tools. This process measured the level of interdisciplinarity, but any value of disparity or interdisciplinarity could not be derived. Along with this delimitation of the study it can be stated that to find the value of interdisciplinarity further study should be curated using some other procedures.

References:

- Evans, E. D. (2016). *Measuring Interdisciplinarity Using Text*. <https://journals.sagepub.com/doi/full/10.1177/2378023116654147>
- Kherwa, P., & Bansal, P. (2018). (PDF) *Topic Modeling: A Comprehensive Review*. https://www.researchgate.net/publication/334667298_Topic_Modeling_A_Comprehensive_Review
- Kim, K., Kogler, D. F., & Maliphol, S. (2024). (PDF) Identifying interdisciplinary emergence in the science of science: Combination of network analysis and BERTopic. *ResearchGate*. <https://doi.org/10.1057/s41599-024-03044-y>
- Laureate, C. D. P., Buntine, W., & Linger, H. (2023). A systematic review of the use of topic models for short text social media analysis. *Artificial Intelligence Review*, 56(12), 14223–14255.
- Paul, M., & Girju, R. (2012). (PDF) *Topic modeling of research fields: An interdisciplinary perspective*. https://www.researchgate.net/publication/228346424_Topic_modeling_of_research_fields_An_interdisciplinary_perspective
- Saha, B. (2021). Application of topic modelling for literature review in management research. *Interdisciplinary Research in Technology and Management*, 249–256.
- Santosa, F. A., Lamba, M., George, C., & Downie, J. S. (2024). Coconut Libtool: Bridging Textual Analysis Gaps for Non-Programmers. *Proceedings of the Association for Information Science and Technology*, 61(1), 639–644. <https://doi.org/10.1002/pra2.1072>
- Singh, S., Singh, S., & Dhir, S. (2023). The evolving relationship of entrepreneurship, technology, and innovation: A topic modeling perspective. *The International Journal of Entrepreneurship and Innovation*, 14657503231179597. <https://doi.org/10.1177/14657503231179597>
- Verma, M. K., & Yuvaraj, M. (2023). (PDF) *AI-Based Literature Reviews: A Topic Modeling Approach*. https://www.researchgate.net/publication/370784698_AI-Based_Literature_Reviews_A_Topic_Modeling_Approach
- Yu, D., & Xiang, B. (2023). Discovering topics and trends in the field of Artificial Intelligence: Using LDA topic modeling. *Expert Systems with Applications*, 225, 120114.