



Strategies for Sustainable Development in Organic Farming: A Topic Modeling Approach Using Latent Dirichlet Allocation

Sumana Chiangnangam ^a, Paiboon Manorom ^{a, b}, Wirapong Chansanam ^{a, *}

^a Khon Kaen University, Khon Kaen, Thailand

^b Bansomdejchaopraya Rajabhat University, Bangkok, Thailand

Abstract

Organic farming products have gained popularity among health-conscious consumers as a sustainable alternative. Agricultural entrepreneurs focusing on organic agriculture aim to create value through sustainable development, leveraging networks across production, processing, and marketing. Despite its growth, organic agriculture development requires further research to identify effective approaches and methods across various dimensions. This study aims to investigate strategies and methodologies for organic farming by analyzing topic models related to organic agriculture and entrepreneurship using bibliographic information. A dataset of 4,327 article abstracts published between 1946 and 2023 in the Scopus database was analyzed using the Latent Dirichlet Allocation (LDA) topic modeling method. The analysis identified five primary topics: (1) Soil for organic agriculture, (2) Environment and organic agriculture, (3) Agricultural business, (4) Organic production, and (5) The use of organic substances. The topic modeling approach yielded a connection value of 0.419, indicating an effective and appropriate grouping of topics. The findings provide valuable insights for farmers, enabling them to understand historical trends and strategies in organic farming and entrepreneurship. Additionally, the study offers researchers a foundation for applying topic modeling to explore future research directions in organic agriculture and related fields.

Keywords: Agricultural entrepreneurs, Latent Dirichlet Allocation, Organic agriculture, Topic Modeling.

1. Introduction

Agriculture in Thailand continuously evolves through technology and has adapted innovation to transform into the era of Agriculture 4.0, which emphasizes high-quality agriculture to increase production productivity in the quantity and value of agricultural products (Office of the Small and Medium Enterprises, 2020). Furthermore, the government sector also expects farmers to practice sustainable agriculture in three dimensions, i.e., the economy, the society, and the environment, to create growth based on an environmentally friendly (Office of the Small and Medium Enterprises, 2021) through promoting agricultural products and processed agricultural products to create high value, which is environmentally friendly and build

*Corresponding author

E-mail addresses: wirach@kku.ac.th (W. Chansanam)

Received: 18 July 2024 Revised: 26 November 2024 Accepted: 27 November 2024

Published: 31 December 2024

an image of Thailand to be tourism destination that emphasizes value and sustainability (National Organic Agriculture Development Committee, 2017).

Organic farming is considered a sustainable agricultural system, which gained the attention of farmers and consumers (Asamoah et al., 2023). This system resulted in the increase of organic agricultural products along with the project for promotion and support for farmer groups to enter organic farming. It also strengthened farmer groups by connecting production, processing, and marketing networks. Notably, in the National Economic and Social Development Plan No. 13 (2023–2027), was mentioned the objectives related to the development of organic agriculture as follows: 1) A high-value economy with a friendly environment and to be a leading country in agricultural products and agricultural processing that create high value and build Thailand to be a tourism destination that emphasizes value and sustainability, 2) Society of opportunity and equality for strong SMEs with high potential and high competitiveness, 3) Sustainable lifestyles for a circular economy and a low-carbon society to reduce risks and impacts from natural disasters and, 4) Factors Driving Development for a highly capable workforce for future development and needs (Announcement on the National Economic and Social Development Plan No. 13., 2022). In addition, the government has urgent policies to lay the foundations of the country's economic system for the future of technology development along with developing the skills of small and medium-sized entrepreneurs (SMEs) and communities as well as urgent policies for developing public services systems to develop the government's data collection and disclosure system to establish a central organic agricultural database for the entire organic agricultural supply chain and support research and development to create systematic knowledge and innovation in a concrete way (Agricultural Research Development Agency, 2018; Office of the Small and Medium Enterprises, 2020).

Before engaging with an entrepreneur, it is essential to gain an in-depth understanding of various critical aspects that can serve as guidelines to drive the growth of entrepreneurial groups. These aspects include information on production systems, production factors, market intelligence, technology, innovation, and other relevant areas. The collection of organic agricultural information should be categorized consistently and rely on updated resources, including information, knowledge, and skills (Blake, Wijetilaka, 2015). Researchers from various fields have been actively studying organic farming and entrepreneurship, with several studies already catalogued in the database. However, finding useful information or additional information about organic agriculture and entrepreneurs makes it difficult to retrieve information in the public sector (Joo et al., 2016). Furthermore, analyzing large amounts of data also requires tools to help in aspect processing information, e.g., natural language processing, which analyzes huge amounts of text from social media data, including topic modeling techniques (Jelodar et al., 2019).

Topic modeling is a data distribution model for categorization and contains ideas from various documents, resulting in a huge amount of data and a collection of topics. Each topic has a probability of words occurring in that topic (Maneewong et al., 2024). Thus, the topic modeling is creating a data distribution model that was used to group data based on the idea that a document is a combination of huge resources. Consequently, the topics might have a probability distribution of many words occurring in each topic based on the concept of the Latent Dirichlet Allocation: LDA, which was created with the idea that a document would consist of topics together randomly in the form of word groups. In the search, LDA is used to find topics or word groups that need to be extracted from the document by calculating the probability value (probabilistic) from words appearing in documents as a latent topic, which cannot be observed. Then, the LDA program will analyze the probability of each word in the latent topic and estimate the proportion of hidden topics in a document. The LDA is the most commonly used, as it is a flexible method and is changeable for creating topic models to separate essential points of the message. To create a topic model, the question can be answered: which topics are discussed most frequently? Therefore, topic modeling will provide more insight into relevant topics within that topic by creating content relevance by selecting semantically relevant keywords. Meanwhile, the content is then analyzed to determine whether it is applicable, which is widely used in various fields of study such as Linguistic science, Political science, Medical and Biomedical Science, Geography, etc. (Jelodar et al., 2019). Hence, this study is an analysis of research on organic agriculture and entrepreneurs in the relevant context and citation, which relevant topics are discussed the most in the field of study.

2. Methods and Materials

This study has employed a quantitative analysis of bibliographic research articles through the text analysis method from research abstracts intending to understand organic agriculture, entrepreneurship, and knowledge pedagogy research trends. Nonetheless, the application of this technique uses the concept of text analysis with a topic model to analyze topic model creation. Then, the researcher considered the prominent words of each issue and grouped them together for the important issues that correspond to the others. The research methodology was comprised of four stages, as were as follows:

Data collection

In this study, the researcher has searched for documents related to organic agriculture and entrepreneurs in the Scopus database by specifying keywords used in the search. To identify the keywords, we then analyzed the main issues regarding organic agriculture and agricultural entrepreneurs in the heading or title of the various sources, such as documents, books, articles, and related research. After that, proceed with the obtained words in the search using advanced search techniques with AND to connect the words to get the results in which both words appear. Besides, the OR was used to connect words to get the results in which either word appears (Prabpala et al., 2023). Therefore, the research technique would be “(organic AND agriculture) OR (organic AND farming) OR (organic AND business) OR (organic AND startup) OR (organic AND entrepreneur)”. The method has selected relevant documents for organic agriculture, entrepreneurs, and businesses. The index organizes the resulting document into several academic and professional fields, including Social Sciences, Business Management and Accounting, Economics/Econometrics and Finance, Multidisciplinary studies, and Arts and Humanities. This classification helps researchers and readers identify the document’s relevance to their area of interest. Additionally, publishers have made the document available exclusively in English, ensuring accessibility to English-speaking audiences. The publication timeline spans from 1946 to 2023, reflecting its long-standing contribution to knowledge in these disciplines. The data in this study was downloaded on April 11, 2023, and found that 6,730 results were obtained. Then, the researcher checked for duplication, anonymous author, and abstract absence to be excluded from this study. The final results were about 4,327 in total and were saved in the file in CSV format for data analysis.

Data preparation and data cleaning with Natural Language Processing (NLP)

For the data preparation and cleaning, the data was imported into English through Natural Language Processing (NLP) at the pre-processing step to clean the data, accelerating computers to understand human language. The NLP would assist in categorizing, summarizing, and creating text, which converts the data with reliable tools as open source and available to run on Python. The text at the pre-processing stage proceeded to PyCaret (Ali, 2020). These steps involved deleting commonly used but insignificant words, which would help to delete the marks, punctuation, numbers, and unnecessary words to extract representative keywords from articles.

Topic Modeling with Latent Dirichlet Allocation

The topic modeling is a tool to help the researcher discover hidden knowledge structures in document datasets. This allows the researcher to make informed decisions and gain insights into complex topics (Blei et al., 2003). However, choosing the right model would be challenging since the various models have different strengths and weaknesses (Jelodar et al., 2019). For example, LDA is well-known for learning descriptive topics, whereas LSA is well-known for generating visual representations of semantics in datasets (Stevens et al., 2012). After completing the pre-processing step, the TF-IDF as weighting method, which is a pre-filtering step for a statistical measure used to rate the importance of words in the content of a document set based on the occurrence of each word, then the relevant keywords in the corpus would be investigated through a bi-gram algorithm to select common phrase together with the TF-IDF algorithm to extract keywords from the abstract (Grun, Hornik, 2011; Wattanasiri et al., 2024).

Data visualization

The data visualization was designed based on group categorizations derived from each topic identified by the Latent Dirichlet Allocation (LDA) model. Graphs and diagrams illustrate the datasets and highlight the relationships within each dimension, providing a clear and structured view of the underlying data patterns.

Data Collection

From collecting and selecting the document lists regarding organic agriculture and entrepreneurs in the Scopus database, the bibliographic information of the articles in each context was recorded, as well as the name of the author, title, year of publication, name of the journal, and issue. The year, the number of citations and the abstract were then downloaded as CSV files, which can be opened in Excel to check and select duplicate data or data without an author. There were no published years. It was found that there were 4,327 articles from 1946 to 2023 (Figure 1). After that, files were prepared for import and analysis, including the name of the author, title, year of publication, and abstract. We added the context for each article in the CSV file format.

3. Results
Published research articles

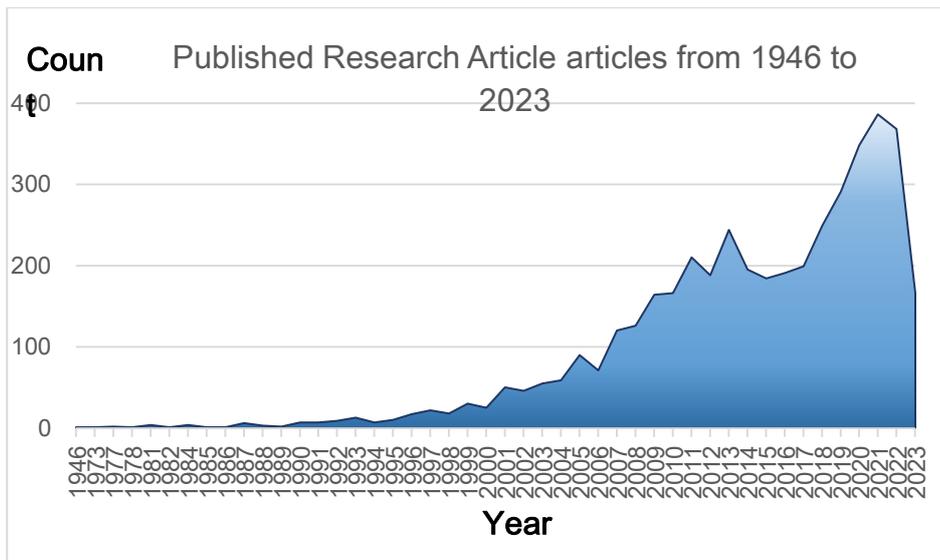


Fig. 1. Published articles from 1976 to 2023

The data reveals that Social Sciences has the highest article concentration, with approximately 1600-1800 publications. This is followed by Business, Management, and Accounting, which show roughly 1000 articles. Arts and Humanities and Economics, Econometrics, and Finance demonstrate similar levels of publication output, each with approximately 600-800 articles. Health Sciences and Multidisciplinary categories show the lowest number of publications, each containing fewer than 400 articles (Figure 2).

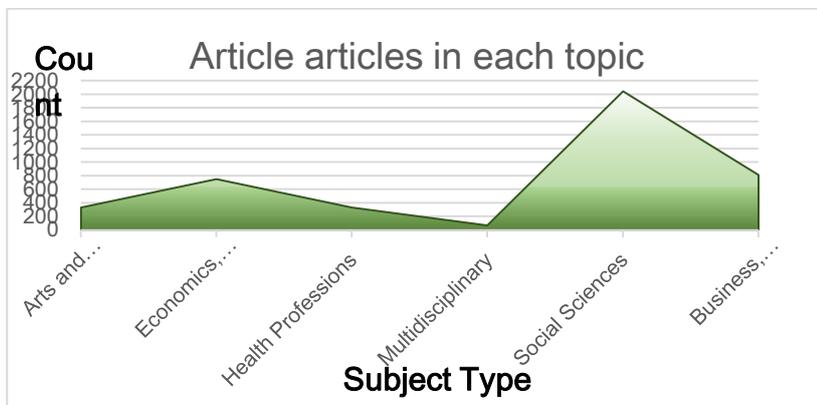


Fig. 2. Article articles on each topic

Data Preparation and Data Cleaning Result

Indicate Data Categorization for Analysis

In this study, we have chosen to analyze the content of the abstract section because the abstract contains an overall summary and contains important information for each article collected by preparing the data and cleaning the data, importing the .CSV file from which the data was collected into the Python program (Figure 3) and examining the data in the abstract section. The results indicated that there were 7,239 session data from 4,327 documents with 19,939 words (Figure 4)

	Authors	Title	Year	Cited by	Abstract	Sbj. A.
0	A.V.V.B.; Banesel I.P.; Weedan G.; Finckh M.B.	Effects of ten years organic and conventional ...	2019	13	Early vigour traits of wheat composite cross p...	0
1	Aaijaz N.; Bin Ibrahim M.D.; Bin Ahmed G.	Green consumers: A growing market for SME'S an...	2010	1	This study attempts to gain knowledge about th...	3
2	Wemink A.J.A.; Hol J.M.G.; Beurskens A.G.C.	Ammonia emission and nutrient load in outdoor ...	2006	16	Ammonia emission and nutrient load in outdoor ...	4
3	Aarts H.F.H.; Conijn J.G.; Corré W.J.	Nitrogen fluxes in the plant component of the ...	2001	14	Sandy areas in th...	4
4	Abbas A.; Sejid M.B.; Sejid J.; Ahmed N.	Forecasting environmental and social benefits ...	2023	1	Embodied carbon of new buildings can be effect...	4
...
4322	Znakić M.; Ječ Rogelj M.; Grgić I.	Organic agricultural production on family farm...	2017	7	The Croatian organic products market is not de...	4
4323	Zuba-Ciszewska M.; Krawaska A.; Manning L.; B...	Organic milk supply in Poland: market and pol...	2019	10	Purpose: Global demand for organic milk produc...	1
4324	Zubizarreta-Gerendiain A.; Pučkola T.; Pellola H.	Effects of wood harvesting and utilisation pol...	2016	30	We studied the effects of different wood harve...	4
4325	Zuo X.X.; Li H.Y.	Carbon sequestration within millet phytoliths ...	2011	91	Phytoliths are noncrystalline minerals that fo...	0
4326	Zuorro A.; Munero-Saber K.A.; González-Delgado...	Evaluating the feasibility of a pilot-scale sh...	2021	6	The foreseen increase in the demand of chitin...	1

Fig. 3. Data from Imported File .CSV format

```
my_nlp_experiment = nlp.setup(data=dataset, target='Abstract')
```

Description	Value
session_id	7239
Documents	4327
Vocab Size	19939

Fig. 4. Document Computation and Appeared Words

Data Processing

(1) First, for a suitable format analysis, cleaning data by eliminating unnecessary data, such as numbers, punctuation marks, symbol marks, and unnecessary words space between words.

(2) After that, the abstract was prepared and imported into the model to analyze topics, which is called topic modeling. We employed Latent Dirichlet Allocation (LDA) to define the topics. Finally, the obtained results from the LDA model showed the measurement values for evaluating topics in Abstract Topic 0 – Topic 4, which is the most valuable topic in the abstract. This was considered the main topic mentioned in the abstract (Figure 5).

(3) In the topic distribution, Figure 3 shows that topic 3 was the most common (Figure 6). Each topic was represented by a group of keywords of the topic in the visualized text in WordCloud format, which is a visual representation of text that highlights keywords in the given content. In this study, we created a word cloud from the WordCloud library in Python to present the main research content and each topic (Figure 7; Cho et al., 2018).

Figure 7 shows a set of word clouds representing different topic clusters related to organic farming and agricultural research.

Topic 0: Centered around “soil” and “use,” with prominent terms like “high” and “crop,” suggesting a focus on soil management and crop cultivation practices.

Topic 1: It prominently features “organic” and “production,” with terms like “cost” and “emission,” indicating research on organic production systems and their environmental/economic impacts.

Topic 2 Emphasizes “use,” “production,” and “environmental,” with “organic” and “system” also prominent. It suggests research on the environmental aspects of organic farming systems.

Topic 3: Dominated by “farmer” and “use,” with “organic” and “farming” also prominent, indicating a focus on farmer practices and adoption of organic farming methods.

Topic 4: It highlights “use,” “study,” and “organic,” with “method” also visible, suggesting a more research-methodology-oriented cluster focusing on organic farming studies.

	Authors	Title	Year	Cited by	Abstract	Sbj. A.	Topic_0	Topic_1	Topic_2	Topic_3	Topic_4	Dominant_Topic	Perc_Dominant_Topic
0	A.V.V.B.; Barisel J.P.; Weedon O.; Finch M.R.	Effects of ten years organic and conventional ...	2019	13	early vigour trait wheat composite cross popul...	0	0.708468	0.119149	0.095792	0.012793	0.063799	Topic 0	0.71
1	Aajez N.; Bin Ibrahim M.D.; Bin Ahmed G.	Green consumers: A growing market for SMEs an...	2010	1	study attempt gain knowledge consumer organic ...	3	0.011254	0.085088	0.183898	0.718953	0.000807	Topic 3	0.72
2	Aernink A.J.A.; Hui J.M.G.; Beurskens A.G.C.	Ammonia emission and nutrient load in outdoor ...	2006	16	ammonia emission load outdoor run lay hen meas...	4	0.546840	0.379436	0.043079	0.028813	0.001831	Topic 0	0.55
3	Aarts H.F.M.; Conijn J.G.; Camé W.J.	Nitrogen fluxes in the plant component of the ...	2001	14	sandy area mainly use intensive dairy farming ...	4	0.558460	0.342258	0.014455	0.083324	0.001503	Topic 0	0.56
4	Abbas A.; Sajid M.B.; Sajid J.; Ahmed N.	Forecasting environmental and social benefits ...	2023	1	embody carbon new building effectively reduce ...	4	0.115941	0.587271	0.112740	0.131617	0.052432	Topic 1	0.59
...
4322	Zrakić M.; Jež Rogelj M.; Grgić I.	Organic agricultural production on family farm...	2017	7	croatian organic product market develop econom...	4	0.008134	0.032815	0.121155	0.835245	0.002650	Topic 3	0.84
4323	Zuba-Ciszewska M.; Kowalska A.; Manning L.; Bt...	Organic milk supply in Poland: market and poll...	2019	10	purpose global demand organic milk product gh...	1	0.003496	0.172735	0.130970	0.648461	0.044338	Topic 3	0.65
4324	Zubizarreta-Genardin A.; Pukkala T.; Peltola H.	Effects of wood harvesting and utilisation pol...	2015	30	study effect different wood harvesting utilisa...	4	0.623315	0.245907	0.006514	0.114429	0.009834	Topic 0	0.62
4325	Zuo X.X.; Lü H.Y.	Carbon sequestration within millet phytoliths ...	2011	91	phytolith noncrystalline mineral form inside c...	0	0.806932	0.013072	0.109274	0.016638	0.054084	Topic 0	0.81
4326	Zuorro A.; Moreno-Sáder K.A.; González-Delgado...	Evaluating the feasibility of a pilot-scale sh...	2021	6	foresee increase demand chitin reveal business...	1	0.060316	0.874597	0.027210	0.018608	0.019268	Topic 1	0.87

4327 rows x 13 columns

Fig. 5. The Results from LDA Model Analysis

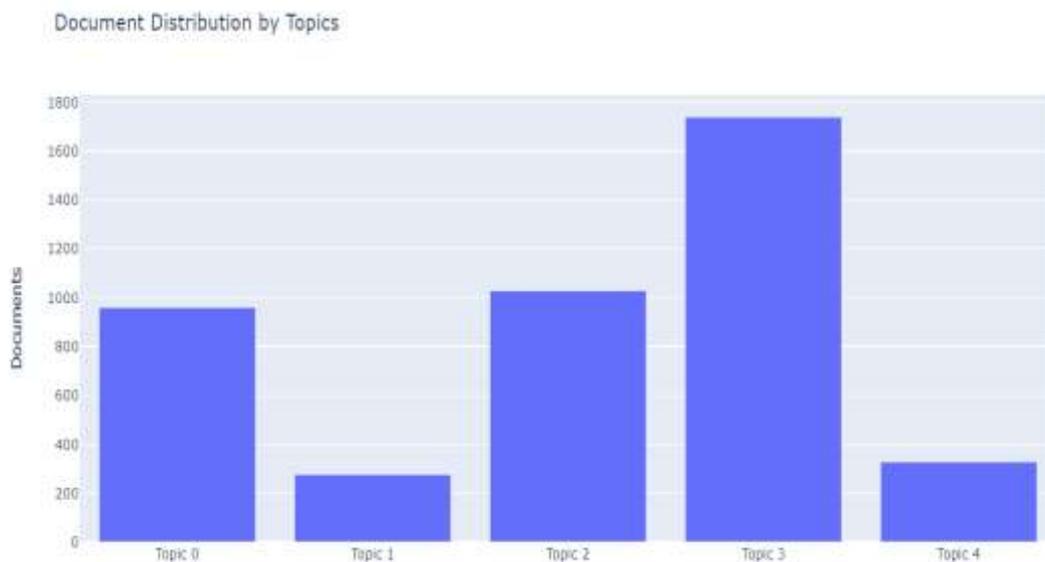


Fig. 6. The Topic Distribution

Common threads across all topics include 1. Consistent appearance of “use” and “organic” across multiple clusters, 2. Environmental and production-related terminology and 3. A mix of practical farming terms and research-oriented language

This WordCloud suggests that these topics represent different aspects of organic farming research, from soil management to farmer practices to production systems. There is some overlap in terminology but distinct thematic focuses.



Fig. 7. WordCloud Format from Library

Topic Modeling Results

Topic modeling is a tool for exploratory analysis of large volumes of documents (Koltsova, Koltcov, 2013; Elgesem et al., 2015), showing the overall interpretability of topics and used to assess topic quality. The Coherence Metrics technique was used to calculate statistical values and probabilities extracted from a reference library, especially focusing on the context of words to score the coherence of topics. From the probability distribution, the model can determine which topics are in a given document and which words are in the given topic by considering the distribution of words on various topics and the distribution of topics in documents. When considering the consistency of topics across topics from 0 to 6, topic 5 (Figure 8) is the optimal number of topics for this model to maximize topic coherence. This corresponds to the 5 topics specified in the pre-processing evaluation.

```

from gensim.models import CoherenceModel
# instantiate topic coherence model
cm = CoherenceModel(model=lda_model, corpus=bow_corpus, texts=corpus, coherence='c_v')
# get topic coherence score
coherence_lda = cm.get_coherence()
print(coherence_lda)

```

0.41970792415122304

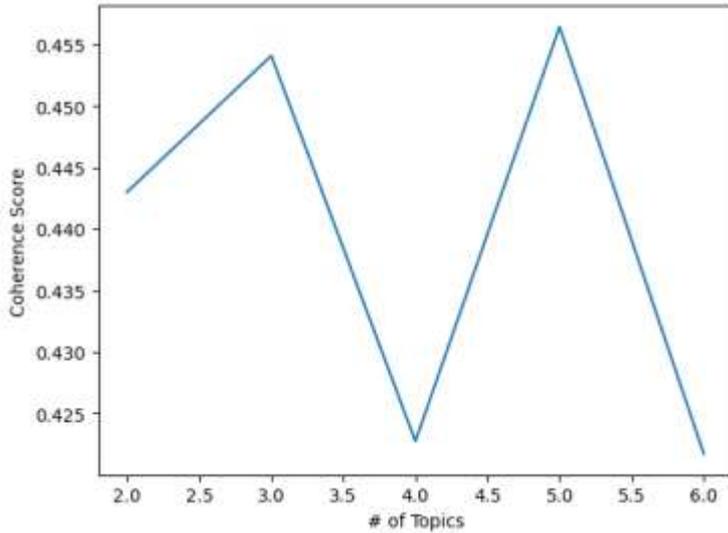


Fig. 8. The Appropriate Topic Computation Results

The Topic Naming and Topic Detail

The researcher has employed the bigram’s properties to predict the most common words in each topic, which the bigram feature performs better for prediction than the unigram and trigram features with overlapping structural relationships in the bibliometric analysis study on the dataset. Among the 100 most common word clusters, five topics were divided according to topic from Topic 0 – Topic 4 (Figures 9–13).

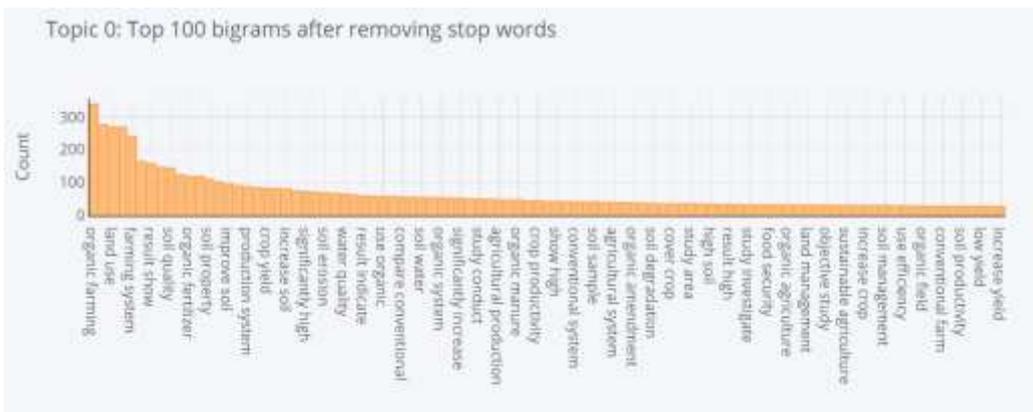


Fig. 9. Word Cluster from Bigram Algorithm in Topic 0

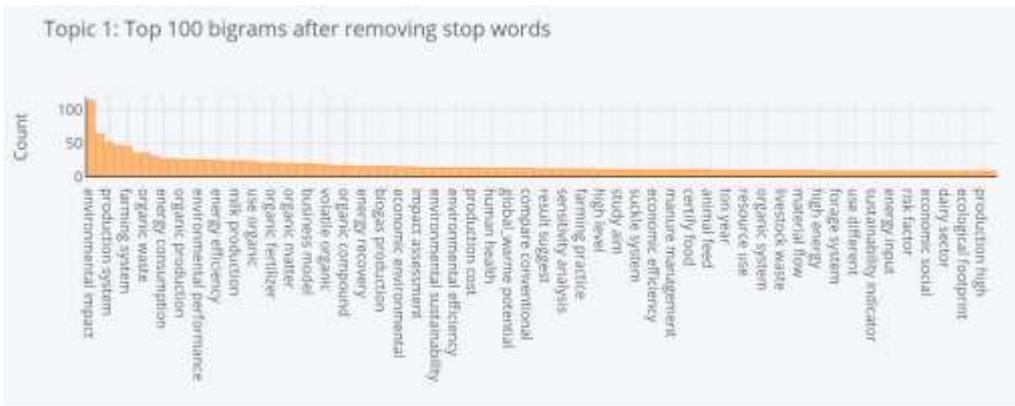


Fig. 10. Word Cluster from Bigram Algorithm in Topic 1

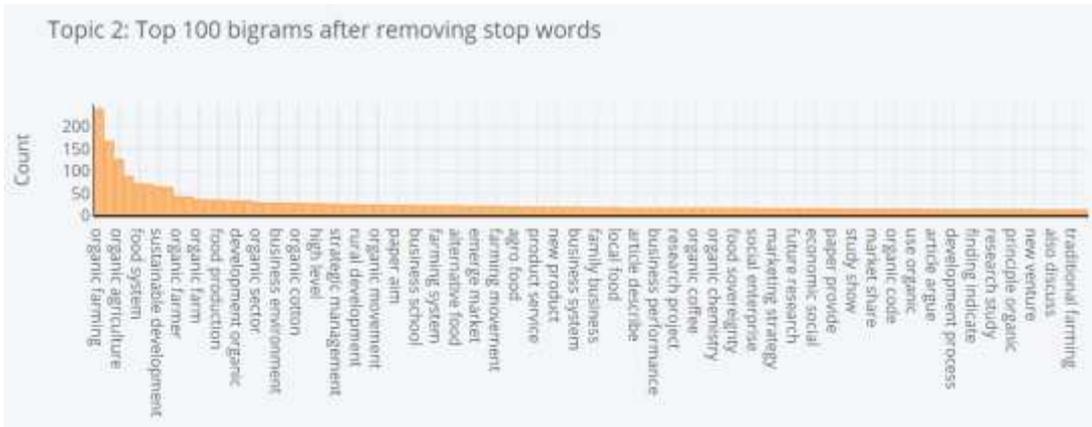


Fig. 11. Word Cluster from Bigram Algorithm in Topic 2



Fig. 12. Word Cluster from Bigram Algorithm in Topic 3

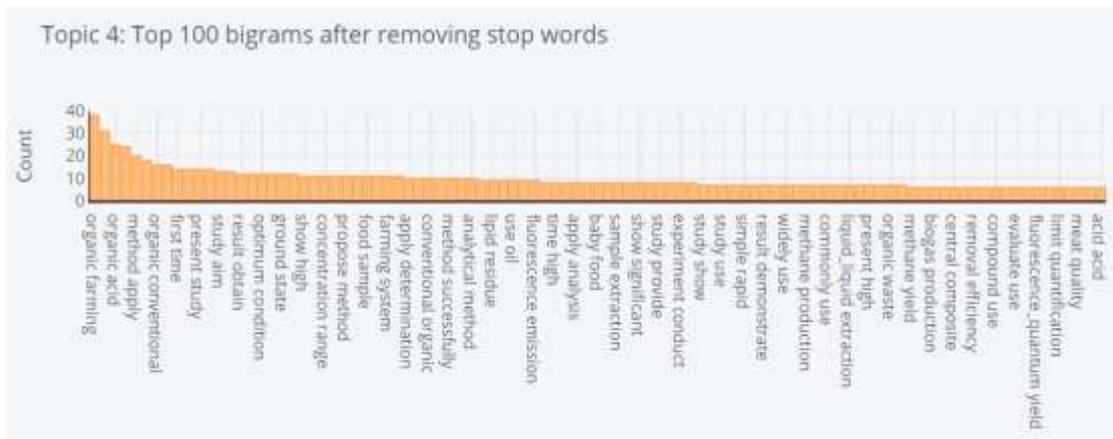


Fig. 13. Word Cluster from Bigram Algorithm in Topic 4

Table 1. The Topic Modeling from the Representative Topic

Topic No.	The Representative Topic (bigrams)	Topic Name	Description
0	Organic farming, organic matter, land use, soil organic, farming system, organic carbon, result show, organic conventional, soil quality, soil fertility, organic fertilizer, cropping system, soil property, farming practice, improve soil, management practice, production system, crop production, crop yield, conventional farming	Soil for Organic Agriculture	The study of organic agriculture requires studying soil in areas related to organic agriculture, such as land use, soil quality, and soil properties. soil improvement adding nutrients to the soil, soil management, etc.
1	Environmental impact, organic farming, production system, result show, farming system, land use, organic waste, case study, energy consumption, management system, organic production, waste management, environmental performance, dairy farm, energy efficiency, energy use, milk production, conventional organic, use organic, organic conventional	Environment and Organics	The origins of organic farming were intended to preserve the environment and reduce its impact on the environment, namely, agriculture using chemicals. This resulted in more than chemical residues in the soil mixed in water and air, continuously affecting consumers' health directly and indirectly. A study of the article shows that this group of liquor information has emerged to have more knowledge and understanding of organic farming.
2	Organic farming, case study, organic agriculture, organic food, food system, organic growth, sustainable development, business model, organic farmer, small business, organic farm, purpose paper, food production, organic production, development	Organic Agriculture Business	Organic farming relates to the business sector. Currently, there are many forms of organic farming business, according to the results of the topic analysis in the article, related to management, production, development, inspection, and strategy, were business systems that classify small businesses, kitchen businesses, new business, including

Topic No.	The Representative Topic (bigrams)	Topic Name	Description
	organic, organic product, organic sector, result show, business environment, social movement		presentation of business results.
3	Organic farming, organic food, organic agriculture, organic product, organic production, organic farm, organic farmer, farming system, result show, organic conventional, food system, case study, farming practice, food product, land use, food production, production system, conventional organic, conventional farming, sustainable agriculture	Organic production	Organic farming consists of many factors to get quality organic products. It maintains organic conditions regarding organic agriculture, organic products, producers, safe food production, food systems, products, organic certification, and agricultural policy.
4	Organic farming, organic solvent, organic acid, result show, method apply, organic conventional extraction method, first time, the result indicated, the present study, volatile organic, study aim, method use, result obtain, organic residue, optimum condition, high quality, ground state successfully apply	Organic use	Having organic farming knowledge about agriculture might not be enough. There is a need to know organic substances so that various compounds can be used correctly, which does not violate the principles of organic farming. The article, therefore, studies and publishes content about organic substances usage and methods for extracting organic compounds, residues, on-organic and organic substances to provide those interested in studying organic agriculture with guidelines for further study.

Topic 0, named “Soil for Organic Agriculture,” was a concept structure that shows the relationship between soil and organic agriculture, which shows the importance of soil used in organic farming is useful in studying. This indicates that studying organic agriculture requires studying soil in areas related to organic agriculture, such as land use, soil quality and properties, soil improvement, adding nutrients to the soil, soil management, etc.

Topic 1, named “Environment and Organics”, was a concept structure that shows the relationship between the environment and organic agriculture, which shows that the origins of organic farming are intended to preserve the environment to reduce the impact on the environment, namely agriculture using chemicals. This resulted in an excess of chemical residues in the soil mixed with water and air, continuously affecting the health of consumers both directly and indirectly. A study of the article shows that this group of liquor information has emerged to have more knowledge and understanding of organic farming.

Topic 2, named “Organic Agriculture Business”, was a concept structure that shows the relationship between business and organic agriculture to manifest how organic farming relates to the business sector. Currently, there are many forms of organic farming businesses, according to the results of the topic analysis in the article, related to management, production, development, inspection, and strategy, which are business systems that classify small businesses, kitchen businesses, new business, including presentation of business results.

Topic 3 is named “Organic production”, a conceptual construct representing organic production. The collected articles revealed that most of the word groups found in this topic cover a

lot of content. Whether it is a farming system, a food system, a production system, an agricultural system, an organic system, an inspection system, organic certification, or sustainable development,

Topic 4, named “Organic use” is a concept structure that expresses the relationship between organic use and methods, which shows content related to inorganic and organic substances, chemical residue, volatile organic, compounds organic solvents, organic acids, compound use, and concentration rate which benefits from this topic and can be used as a guideline for further studies regarding organic agriculture apart from agriculture or farming.

Data Visualization

Interactive diagramming is a highly effective way to present the results of topic models. In this regard, the pyLDAVis package (Sievert, Shirley, 2014) created interactive diagrams showing the most representative topics and terms. (Figure 14). The size of each circle in the diagram indicates the relevance of the topic in the corpus, and topics that are close together are more similar. One of the main advantages of the pyLDAvis visualization method is that users can adjust the relevance of words in a topic using a slider (Chuang et al., 2012; Sievert, Shirley, 2014). This tool offers clear and intuitive visualization and can show the relationships and strengths of each topic by displaying the words that make up each topic in a circle and horizontal bar chart.

The circles on the left panel show the model’s overall view, allowing users to easily understand the relationship between topics and their related strengths. Meanwhile, the horizontal bar chart on the right panel shows the terms that make up each topic, which gives the user a detailed understanding of the topics.

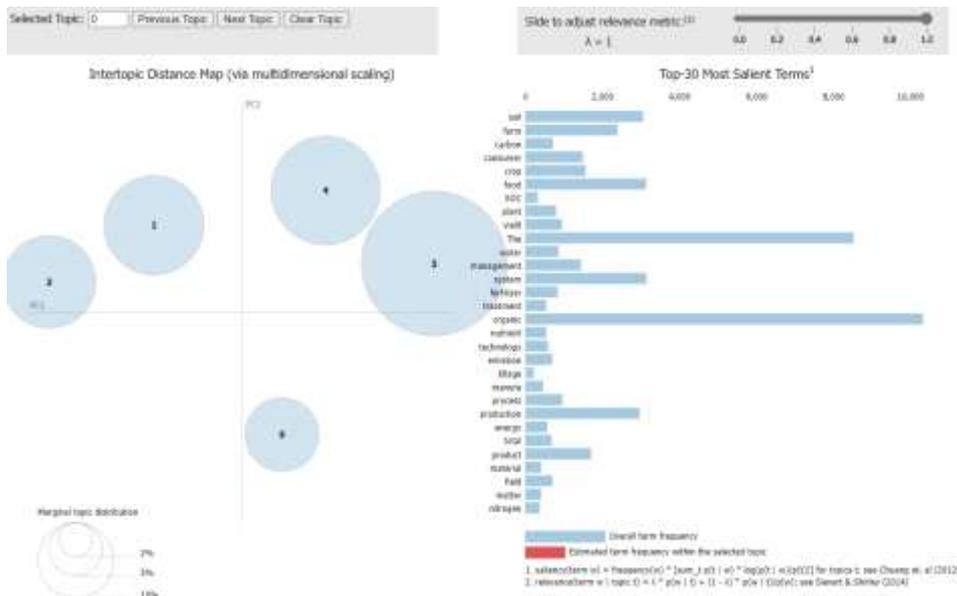


Fig. 14. Show pyDAVis topic model results.

Source: [Ischool](#)

In this study, it can be clearly seen that the five topics identified were different and belonged to different research areas. When clicking on each topic circle, the tool creates a bar graph showing the top 30 for the most relevant terms for that topic. This feature allows users to quickly and concisely summarize topic relevance through the most important keywords by analyzing the words of these keywords to make it possible to categorize all five topics.

The categories covered by these topics were highly relevant to current research topics. This can be seen from the keywords related to each topic by plotting document classifications using t-SNE (t-distributed Stochastic Neighbor Embedding). In this method, each group was represented as a probability distribution, which was a normal distribution to measure the distance between groups. This technique was used for 3D projections to visualize the similarity between multidimensional vectors and plot groups of similar documents (Figure 15).



Fig. 15. The Topic Distribution representing in 3-Dimensions

4. Discussion

This study offers a detailed examination of organic agriculture and entrepreneurship by analyzing bibliographic data from 4,327 articles published between 1946 and 2023, sourced from the Scopus database. By employing the Latent Dirichlet Allocation (LDA) method, the research identified five significant topics that provide essential insights into the multifaceted nature of organic agriculture and its connection with environmental sustainability, economic development, and technological innovation. These topics include soil management for organic agriculture, the interplay between organic farming and environmental conservation, the integration of organic farming into business models, organic production systems, and the effective use of organic substances.

The findings highlight the importance of soil as the foundation for sustainable organic farming, emphasizing soil quality, management practices, and nutrient enhancement. This perspective aligns with existing literature, such as that of Blake and Wijetilaka (2015), which underscores the critical role of soil in organic systems. Furthermore, research focused on reducing chemical residues and promoting environmental sustainability through organic farming resonates with the priorities outlined in the National Organic Agriculture Development Strategy (2017-2021). This strategy emphasizes environmentally friendly agricultural practices as a cornerstone of organic farming.

The study also illuminates the relationship between organic farming and the business sector, illustrating how organic agriculture has been integrated into small enterprises and supply chain networks. These findings support the goals of the National Economic and Social Development Plan No. 13 (2023–2027), which stresses economic growth and entrepreneurship in sustainable agriculture. Similarly, the results underscore the necessity for quality assurance, organic certification, and sustainable food systems, aligning with global trends emphasized by the Research Institute of Organic Agriculture (2021). In addition to exploring soil, environment, and business aspects, the study identifies the methods and standards for using organic compounds effectively and safely. As Joo et al. (2016) noted, this focus aligns with modern agricultural innovations, ensuring adherence to organic farming principles while promoting sustainable practices.

The study's findings validate and expand upon the claims presented in the introduction. It reinforces the evolving nature of organic agriculture, particularly its significance in achieving sustainable development and economic progress. The introduction highlighted Thailand's Agriculture 4.0 initiative, which integrates technology, innovation, and sustainability into agricultural practices (Office of the Small..., 2020). This research offers empirical support for these

objectives, demonstrating how advanced topic modeling and visualization tools such as pyLDAVis can structure diverse research findings into coherent topics (Chansanam, Tuamsuk, 2020). Moreover, the study addresses the challenges noted in the introduction, such as the difficulty in retrieving organized information about organic farming and entrepreneurship (Joo et al., 2016), by providing a systematic and accessible framework.

This study's implications are far-reaching, as it bridges the gap between theoretical research and practical applications. By visualizing trends in organic agriculture, the research offers a clear roadmap for farmers, policymakers, and researchers, enabling them to identify key areas for focus and development. For example, insights into soil management practices and the use of organic substances directly benefit farmers, while understanding business models for organic agriculture informs policymakers and entrepreneurs. Despite its strengths, the study has limitations. Its reliance on abstract data may overlook detailed findings in full-text articles, and its focus on English-language publications may exclude significant contributions in other languages. Future research should expand to include broader datasets and explore additional scientific fields to create a more comprehensive understanding of organic agriculture.

In conclusion, this study enhances the understanding of organic agriculture and entrepreneurship by systematically categorizing a diverse body of research. The insights gained lay a solid foundation for future investigations and practical applications, significantly contributing to the global sustainable development of organic farming practices.

5. Conclusion

The analysis of research on organic agriculture and entrepreneurs from the barometric data was done using the text analysis technique with the PyCaret library in Python, and a group of words related to organic agriculture and agricultural producers were obtained. The results have expanded to reveal more knowledge, which is useful for further study in the parts that are not yet complete to be able to use it to structure knowledge more comprehensively. Especially, issues regarding the content and important concepts of organic agriculture and entrepreneurs. There were issues that farmers or entrepreneurs must study, from organic farming, production systems, development, management, and marketing to knowledge about the environment and essential organic substances to be able to do organic farming correctly according to standards and have sufficient knowledge in doing organic farming business.

This study analyzed the bibliographic data using text analysis and topic modeling techniques on abstracts from 4,327 articles collected from the Scopus database, analyzing five different aspects. The Python was employed to perform topic modeling and evaluate each topic's performance. The model and present figure of the results clearly show the information, emphasizing insights and important trends from the analysis.

In the aspect of limitations in this study, this study reflected in the analysis section that the researcher had to choose stable programs and software that were able to analyze and process accurately and precisely, and that could be recorded and stored appropriately. In addition, the results of this study were a collection of information in the context related to organic agriculture and entrepreneurs, which was analyzed from the abstract. There might be some limits to reading research results that might not cover all of agricultural science. Therefore, future research should explore additional fields of science to get more comprehensive results.

6. Strengths and Limitations

A major strength of the study is its robust methodology and use of advanced tools like LDA and pyLDAVis. These tools effectively categorize and visualize complex data, making it accessible to farmers, policymakers, and researchers. The study aligns well with global and national strategies, such as Thailand's Agriculture 4.0, emphasizing its practical relevance.

However, its reliance on abstracts may overlook deeper insights in full texts, and its focus on English-language articles may exclude significant regional contributions. Additionally, while topics are well-defined, their interrelations are not fully explored. Despite these limitations, the study provides a strong foundation for future research and practical applications in sustainable organic farming.

7. Implications of the Study

The findings of this study have significant implications for the development of organic agriculture and entrepreneurship. By identifying five core topics – soil management, environmental impacts, organic business models, production systems, and organic substance usage – the study provides a structured framework to guide stakeholders in addressing key challenges in sustainable farming. Farmers can leverage insights on soil quality and sustainable practices to enhance productivity while adhering to organic standards. Policymakers can use the findings to design targeted strategies that promote organic agriculture, aligning with broader goals such as environmental sustainability and economic growth, as seen in initiatives like Thailand's Agriculture 4.0.

The study offers a methodological blueprint for researchers, showcasing the utility of advanced tools like LDA in extracting meaningful insights from large datasets. This facilitates the exploration of emerging trends and gaps in organic farming research. Overall, the study bridges theoretical knowledge with practical applications, fostering innovation and sustainable practices in organic agriculture.

8. Declarations

Ethics approval and consent to participate

This study did not involve human participants, animals, or any sensitive personal data and therefore did not require approval from an Institutional Review Board (IRB) or ethics committee. The research was conducted using publicly available bibliographic data from the Scopus database, adhering to all applicable guidelines for ethical research. No consent to participate was required as the study relied solely on secondary data sources without the direct involvement of individuals or organizations.

Consent for publication

All authors have reviewed and approved the final manuscript for submission. They consent to the publication of this work and confirm that it has not been submitted elsewhere for publication. Each author agrees to be accountable for all aspects of the work, ensuring the integrity and accuracy of the study.

Availability of data and materials

Available upon formal request to the corresponding author.

Conflict of interest statement

The authors report no conflicts of interest.

Funding

This research received no external funding. However, the authors sincerely thank the Centre for Behaviour and Wellness Advocacy, Ghana, for providing financial support through the Institutional Open Access Publication Fund to automatically provide a full waiver for the article processing fee.

Authors' contributions

All authors (S.C., P.M., and W.C.) contributed equally to this work. They collaboratively developed the concept and design of the study, collected data, and contributed to the analysis and interpretation of the collected data. Additionally, all authors were involved in drafting, revising, and finalizing the manuscript.

Acknowledgements

This research is conducted under a cooperative agreement between the Centre for Behaviour and Wellness Advocacy, Ghana, and the Faculty of Humanities and Social Sciences at Khon Kaen University, Thailand. This work has received a scholarship under the Research Assistant Program from Khon Kaen University, Thailand (Grant No. RA2566-12).

Authors' ORCID

Sumana Chiangnangam  <https://orcid.org/0009-0003-3516-681X>

Paiboon Manorom  <https://orcid.org/0009-0002-2165-1120>

Wirapong Chansanam  <https://orcid.org/0000-0001-5546-8485>

References

Agricultural Research Development Agency, 2018 – Agricultural Research Development Agency. (2018). Modern agriculture has transformed into a full-fledged businessman. 2018.

[Electronic resource]. URL: https://www.arda.or.th/knowledge_detail.php?id=24 (date of access: 20.11.2023).

Ali, 2020 – Ali, M. (2020). PyCaret – An open source low-code machine learning library. [Electronic resource]. URL: <https://www.linkedin.com/pulse/pycaret-open-source-low-code-machine-learning-ali-mma-cpa-cma> (date of access: 20.08.2023).

Announcement..., 2022 – Announcement on the National Economic and Social Development Plan No. 13. (2022, 24, October). Royal Gazette, Volume 139, Special Section 258D.

Asamoah et al., 2023 – Asamoah, W.E.F., Appiah, P., Tham-Agyekum, E.K. (2023). Assessment of the usage of extension communication channels for disseminating crop production information to oil palm farmers in the Birim South District, Ghana. *Journal of Advocacy, Research and Education*. 10(3): 133-144.

Asmussen, Møller, 2019 – Asmussen, C.B., Møller, C. (2019). Smart literature review: A practical topic modelling approach to exploratory literature review. *Journal of Big Data*. 6(1): 1-18. DOI: <https://doi.org/10.1186/s40537-019-0255-7>

Bettina, Kurt, 2011 – Bettina, G., Kurt, H. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*. 40(13): 1-30. DOI: <https://doi.org/10.18637/jss.v040.i13>

Blake, Wijetilaka, 2015 – Blake, M., Wijetilaka, S. (2015). 5 tips to grow your startup using SWOT analysis. [Electronic resource]. URL: <http://www.afr.com/it-pro/5-tips-to-grow-your-startup-usingswot-analysis-20150226-13pkj5>. (date of access: 20.08.2023).

Blei et al., 2003 – Blei, D.M., Ng, A.Y., Jordan, M.I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*. 3(Jan): 993-1022.

Chansanam, Tuamsuk, 2020 – Chansanam, W., Tuamsuk, K. (2020). Thai Twitter sentiment analysis: Performance monitoring of politics in Thailand using text mining techniques. *International Journal of Innovation, Creativity and Change*. 11(12): 436-452.

Cho et al., 2018 – Cho, S.B., Shin, S., Kang, D.S. (2018). A study on the research trends on open innovation using topic modeling. *Informatization policy*. 25(3): 52-74. DOI: <https://doi.org/10.22693/NIAIP.2018.25.3.052>

Chuang et al., 2012 – Chuang, J., Manning, C.D., Heer, J. (2012). Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the international working conference on advanced visual interfaces* (pp. 74-77).

Diego et al., 2020 – Diego, F.B. Mario, G., David, G., Sergio, M.L. (2020). Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach. *IEEE Access*. 8: 35318-35330. DOI: <https://doi.org/10.1109/ACCESS.2020.2974983>

Donthu et al., 2021 – Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., Lim, W.M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*. 133: 285-296. DOI: <https://doi.org/10.1016/j.jbusres.2021.04.070>

Elgesem et al., 2015 – Elgesem, D., Steskal, L., Diakopoulos, N. (2019). Structure and content of the discourse on climate change in the blogosphere: The big picture. In *climate change communication and the internet* (pp. 21-40). Routledge.

Grun, Hornik, 2011 – Grun, B., Hornik, K. (2011). topicmodels: an R package for fitting topic models. *Journal of Statistical Software*. 40(13): 1-30. DOI: <https://doi.org/10.18637/jss.v040.i13>

Ischool – Ischool. [Electronic resource]. URL: https://ischool.kku.ac.th/bibliometric/lda_organic_agriculture.html (date of access: 20.11.2023).

Jelodar et al., 2019 – Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*. 78: 15169-15211. DOI: <https://doi.org/10.1007/s11042-018-6894-4>

Joo et al., 2016 – Joo, S., Koide, S., Takeda, H., Horyu, D., Takezaki, A., Yoshida, T. (2016). Agriculture activity ontology: an ontology for core vocabulary of agriculture activity. In *International Semantic Web Conference (Posters & Demos)*, pp. 320-335.

Koltsova, Koltcov, 2013 – Koltsova, O., Koltcov, S. (2013). Mapping the public agenda with topic modeling: The case of the Russian Live Journal. *Policy Internet*. 5(2): 207-227. DOI: <https://doi.org/10.1002/1944-2866.POI331>

Li, Lei, 2021 – Li, X., Lei, L. (2021). A bibliometric analysis of topic modeling studies (2000–2017). *Journal of Information Science*. 47(2): 161-175. DOI: <https://doi.org/10.1177/0165551519877049>

[Maneewong et al., 2024](#) – Maneewong, C., Boonwanna, T., Nguyen, L. T., Li, C., Chansanam, W. (2024). The Relationship between China and Vietnam in Research Documents: Text Analytics. *International Research Journal of Multidisciplinary Scope*. 5(4): 98-116. DOI: <https://doi.org/10.47857/irjms.2024.v05i04.01181>

[National Organic Agriculture Development Committee, 2017](#) – National Organic Agriculture Development Committee. (2017). National Organic Agriculture Development Strategy (2017-2021). Bangkok: Office of Agricultural Economics.

[Office of the Small and Medium Enterprises, 2020](#) – Office of the Small and Medium Enterprises. (2020). Report of Strategy and action plan for promoting small and medium enterprises in the agricultural industry. Office the Small and Medium Enterprises. 2020. [Electronic resource]. URL: https://www.sme.go.th/upload/mod_download/download-20191022060242.pdf (date of access: 20.11.2023).

[Prabpala et al., 2023](#) – Prabpala, S., Chansanam, W., Tuamsuk, K. (2023). Topic Modeling on Crowd Trading Ideas for Digital Asset Price Prediction. *International journal of Applied Engineering and Technology (London)*. 5(1): 6-12.

[Research Institute of Organic Agriculture, 2021](#) – Research Institute of Organic Agriculture. (2021). Report of The World of Organic Agriculture 2021. [Electronic resource]. URL: http://www.pad.moi.go.th/images/form_download/.pdf (date of access: 20.11.2023).

[Sievert, Shirley, 2014](#) – Sievert, C., Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces. (pp. 63-70)

[Wattanasiri et al., 2024](#) – Wattanasiri, P., Manorom, P., Chansanam, W. (2024). Analyzing global trends and collaborations in library and information science research: A bibliometric and social network analysis perspective. *International Research Journal of Multidisciplinary Scope*. 5(4): 117-130. DOI: <https://doi.org/10.47857/irjms.2024.v05i04.01229>