Analyzing Research Trends in Organic Agriculture and Entrepreneurship: A Topic Modeling Approach Using Latent Dirichlet Allocation (LDA)

Sumana Chiangnangam[1], Paiboon Manorom[1,2], Wirapong Chansanam[1*]

[1]Faculty of Humanities and Social Sciences, KhonKaen University, Thailand
[2]Faculty of Humanities and Social Sciences, BansomdejchaoprayaRajabhat University, Bankok, Thailand
Email: wirach@kku.ac.th
*Corresponding Author

Received: 11.04.2024 Revised : 14.05.2024 Accepted: 29.05.2024

ABSTRACT
Organic farming products are becoming popular among healthy consumers as an alternative consumption method. Agricultural entrepreneurs interested in developing organic agriculture have emphasized organic farming to create value based on sustainable development through networking in production, processing, and marketing. Nevertheless, the development of organic agriculture still requires support studies to find appropriate approaches and methods to develop organic agriculture in various aspects. Thus, to investigate the appropriate techniques and strategies for organic farming, this research aims to analyze the topic model related to organic agriculture and entrepreneurship from bibliographic information by analyzing the abstract methods for 4,327 published articles from 1946 to 2023 in the Scopus database by using the Latent Dirichlet Allocation (LDA) topic modeling method. The study results, in this study, indicated that the five different topics were as follows: 1) Soil for organic agriculture, 2) Environment and Organic agriculture, 3) Agriculture Business, 4) Organic production, and 5) The Use of organic substances. This study contained a connection value of 0.419, which indicated that the topic modeling was at the appropriate measurement for grouping model topics. Therefore, the topics modeling in this study would help the farmers to notice the ideas related to organic farming and entrepreneurship from the past and assist the researcher in applying the research topics and modeling for future research.

Keywords: Organic agriculture, Agricultural entrepreneurs, Topic modeling, Latent Dirichlet Allocation, LDA

1. INTRODUCTION

Agriculture in Thailand is evolving by technology continuously and has adapted innovation to transform into the era of Agriculture 4.0, which emphasizes high-quality agriculture to increase production productivity in quantity and value of agricultural products [1]. 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 [2] through promoting agricultural products and processed agricultural products to create high value, which is environmentally friendly and build an image of Thailand to be tourism destination that emphasizes value and sustainability [3].

Organic farming is considered a sustainable agricultural system, which gained a number of attention from farmers and consumers and resulted in the organic agricultural products have increased along with the project for promotion and support for farmer groups to enter organic farming and strengthen farmer groups by connecting networks in production, processing, and marketing. Particularly, 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) High-value economy with 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 [4]. 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 develop 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 [5], [1].

Before getting to know the entrepreneur, there are numerous information that might need to be understood in depth to be used as a guideline to drive the growth of entrepreneurial groups, such as production system information, production factors, market intelligence, technology, innovation information, and so on. Due to the collection of organic agricultural information should be categorized in the same way and rely on the updated resources of information, knowledge, and skills [6], with researchers in various fields conducting research on organic farming and entrepreneurship and there were a number of studies have listed in the database. However, finding useful information or additional information about organic agriculture and entrepreneurs seems to be difficult to retrieve information into the public sector [8]. 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 [9].

Topic modeling is a model of data distribution 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. Thus, the topic modeling is the creation of a data distribution model which was used in grouping 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 aspect of searching, 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 clearly 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 which comprises of a flexible method and changeable for creating topic models to separate important 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 through selecting semantically relevant keywords. Meanwhile, the content is then analyzed to determine whether it is relevant or not, which is widely used in various fields of study such as Linguistic science, Political science, Medical and Biomedical, Geographical and locations, etc. [9]. 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. METHOD

This study has employed a quantitative analysis of bibliographic research articles through the text analysis method from research abstracts with the aim to understand the research trends of organic agriculture, entrepreneurship, and knowledge pedagogy. 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:

2.1 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. 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, to identify the keywords. 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 [10]. 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 to organic agriculture, entrepreneurs, and businesses. Thus, the appeared document would be indexed in the fields of Social Sciences, Business Management and Accounting, Economics/Econometrics and Finance, Multidisciplinary, Arts and Humanities and has been published from 1946 to 2023 in English only. 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 as aCSV format for data analysis.

2.2 Data Preparation and Data Cleaning with Natural Language Processing (NLP)

For the data preparation and cleaning, the data was imported in the English language through the Natural Language Processing (NLP) at the pre-processing step to clean the data, which accelerates 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 proceeded in PyCaret [11]. 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.

2.3 Topic Modelling with Latent Dirichlet Allocation

The topic modelling is a tool to help the researcher discover hidden knowledge structures in document datasets. This allows the researcher to be informed decisions and gain insights into complex topics [12]. However, choosing the right model would be challenging since the various models have different strengths and weaknesses [9]. For example, LDA is well-known for descriptive topics learning, whereas the LSA is well-known for generating visual representations of semantic in datasets [13]. 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 [14].

2.4 Data Visualization

The data visualization was divined based on the group categorization in each topic from the LDA in the graph and diagram to illustrate the datasets and relationships in each dimension.

3. RESULTS

3.1 Data Collection Result

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, number of citations, and abstract then downloaded as CSV file, 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. After that, files were prepared for import and analysis, including the name of the author, title, year of publication, and abstract, and added context for each article with the CSV file format, which can be imported into topic analysis programs.

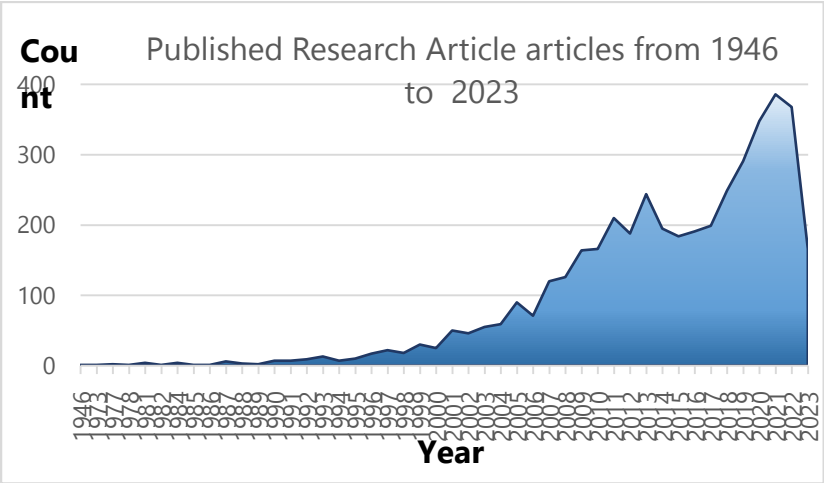

Figure 1. Published Research Article articles from 1946 to 2023

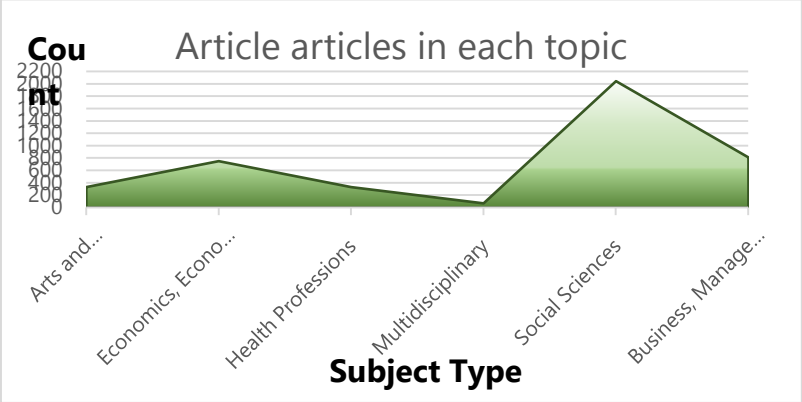

Figure 2. Article articles in each topic

3.2 Data Preparation and Data Cleaning Result

3.2.1 Indicate Data Categorization for Analysis

In this study, the researcher has 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)

```
display(dataset)
```

	Authors	Title	Year	Cited by	Abstract	Sbj. A.
0	A.V V.B.; Baresel J.P.; Weedon O.; Finckh M.R.	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	Aarnink 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.M.; Conijn J.G.; Corré W.J.	Nitrogen fluxes in the plant component of the ...	2001	14	Sandy areas in th...	4
4	Abbas A.; Sajid M.B.; Sajid J.; Ahmed N.	Forecasting environmental and social benefits ...	2023	1	Embodied carbon of new buildings can be effect...	4
...	...	...	...	...	...	...
4322	Zrakić M.; Jež 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.; Kowalska A.; Manning L.; Br...	Organic milk supply in Poland: market and poli...	2019	10	Purpose: Global demand for organic milk produc...	1
4324	Zubizarreta-Gerendiain A.; Pukkala T.; Peltola H.	Effects of wood harvesting and utilisation pol...	2016	30	We studied the effects of different wood harve...	4
4325	Zuo X.X.; Lü H.Y.	Carbon sequestration within millet phytoliths ...	2011	91	Phytoliths are noncrystalline minerals that fo...	0
4326	Zuorro A.; Moreno-Sader K.A.; González-Delgado...	Evaluating the feasibility of a pilot-scale sh...	2021	6	The foreseen increase in the demand of chitin,...	1

4327 rows × 6 columns

Figure 3. Data from ImportedFile .CSV format

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

Description	Value
session_id	7239
Documents	4327
Vocab Size	19939

Figure 4. Document Computation and Appeared Words

3.2.2 Data Processing

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

(2) After that, the abstract has been prepared and imported into the model to analyze topics called topic modelling. Then, employed Latent Dirichlet Allocation (LDA) to define the topics. Finally, the obtained results from the LDA model show 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).

	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.; Baresel J.P.; Weedon O.; Finckh 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	Aaijaz N.; Bin Ibrahim M.D.; Bin Ahmed G.	Green consumers: A growing market for SME'S 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	Aarnink A.J.A.; Hol 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.; Corré 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.; Br...	Organic milk supply in Poland: market and poli...	2019	10	purpose global demand organic milk product giv...	1	0.003496	0.172735	0.130970	0.648461	0.044338	Topic 3	0.65
4324	Zubizarreta-Gerendiain A.; Pukkala T.; Peltola H.	Effects of wood harvesting and utilisation pol...	2016	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-Sader 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 × 13 columns

Figure 5. The Results fromLDA Model Analysis

(3) In the topic distribution, the figure shows that topic 3 was the most common topic (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. The Wordcloud, in this study, was created from the WordCloud library in Python to present the main research content and each topic (Figure 7) [15].

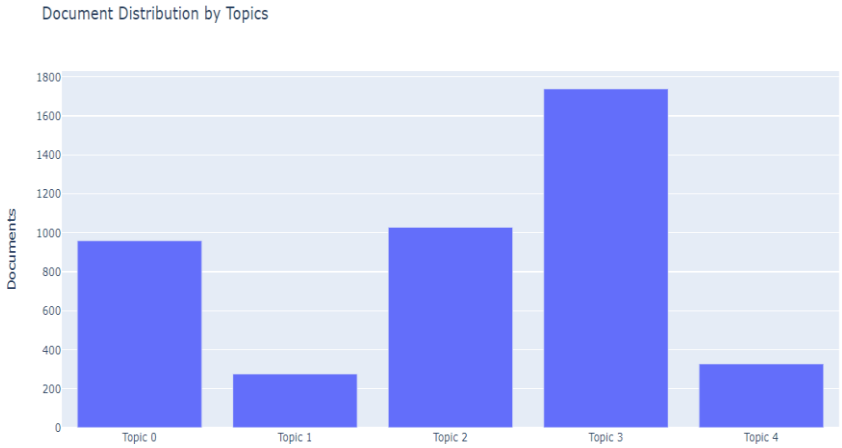

Figure 6. The Topic Distribution

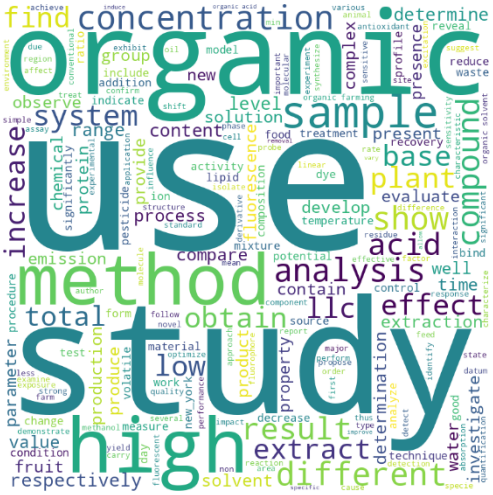

Figure 7. WordCloud Format from Library

3.3 Topic Modeling Results

3.3.1 Topic modelling is a tool for exploratory analysis of large volumes of documents [16], [17], 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), the optimal number of topics to use 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
```

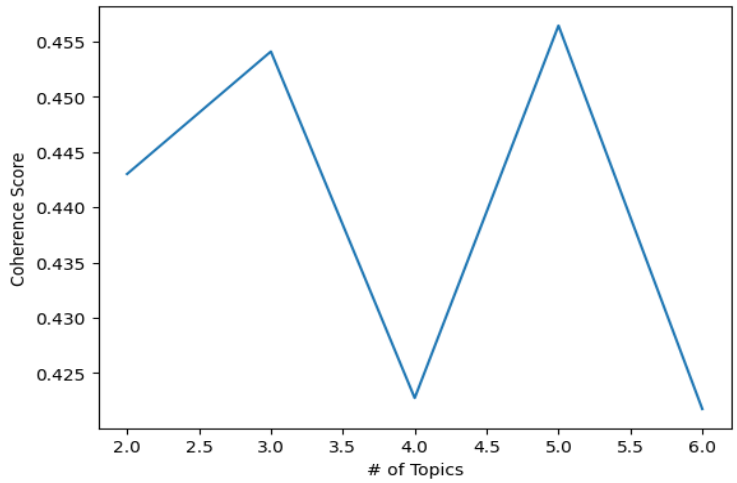

Figure 8 . The Appropriate Topic Computation Results

3.3.2 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 (Figure 9 - 13).

Figure 9. Word Cluster from Bigram Algorithm in Topic 0

Figure 10. Word Cluster from Bigram Algorithm in Topic 1

Figure 11. Word Cluster from Bigram Algorithm in Topic 2

Figure 12. Word Cluster from Bigram Algorithm in Topic 3

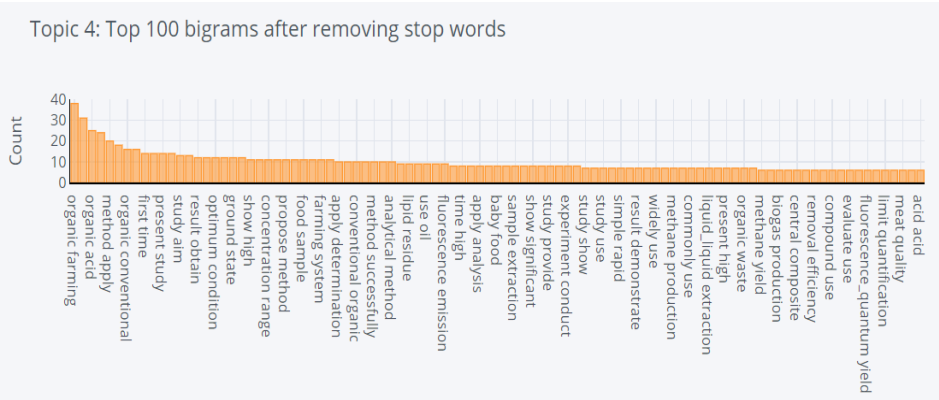

Figure 13. Word Cluster from Bigram Algorithm in Topic 4

Table 1. The Topic Modelling 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 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 to reduce the impact on the environment, namely agriculture using chemicals. This resulted more than chemical residues in the soil mixed in 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.
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, organic product, organic sector, result show, business environment, social movement	Organic Agriculture Business	The 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 business, new business, including 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	The organic farming has consisted of many factors to get quality organic products and maintains organic condition about organic agriculture, organic products, organic producer, Safe-food production, food systems, food products, including organic certification and agricultural policy.
4	Organic farming, organic solvent, organic acid, result show, method apply, organic conventional extraction method, first time, result indicate, present study, volatile organic, study aim, method use, result obtain,	Organic use	Having the organic farming knowledge about agriculture might not enough but may need to have knowledge about organic substance to be able to use various compounds correctly, which does not violate the principles of organic farming. The article therefore studies and

Topic No.	the Representative Topic(bigrams)	Topic Name	Description
	organic residue, optimum condition, high quality, ground state successfully apply		publishes content about organic substances usage, methods for extracting organic compounds, residues, on-organic and organic 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 shows that studying 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.

Topic 1: "Environment and Organics"

The concept structure of Topic 1, titled "Environment and Organics," elucidates the intricate relationship between the environment and organic agriculture. This structure highlights that the origins of organic farming are deeply rooted in the intention to preserve and protect the environment. Organic farming practices emerged as a response to the detrimental impacts of conventional agriculture, which heavily relies on chemical inputs. The excessive use of chemicals in traditional farming leads to significant environmental degradation, including accumulating chemical residues in the soil, contamination of water sources, and air pollution. These pollutants have a cascading effect, impacting the immediate ecosystem and posing direct and indirect health risks to consumers. Organic agriculture aims to mitigate these adverse effects by promoting environmentally sustainable and health-conscious farming practices. The transition to organic farming involves adopting techniques that enhance soil health, conserve water, reduce pollution, and promote biodiversity. By eliminating synthetic chemicals, organic farming reduces the risk of chemical residues entering the food chain, safeguarding consumer health. A comprehensive study of the relevant literature indicates a growing body of knowledge and understanding within the community regarding the benefits and practices of organic farming. This increased awareness is critical for the continued advancement and adoption of organic agriculture, as it underscores the importance of sustainable practices for environmental preservation and public health. The findings suggest an emerging trend towards greater recognition of organic farming's role in addressing environmental challenges and promoting a healthier, more sustainable food system.

Topic 2: "Organic Agriculture Business"

The "Organic Agriculture Business" concept provides a comprehensive framework illustrating the complex relationships between business and organic farming. This framework encompasses key domains including management, production, development, inspection, and strategy, each representing crucial aspects of organic farming enterprises. The organic agriculture business landscape is diverse, ranging from small family-owned farms to innovative new ventures. These businesses are categorized based on scale and focus, including small local operations, kitchen businesses integrating farming with culinary activities, and emerging enterprises bringing innovation to the sector. The framework emphasizes the importance of effective management practices, sustainable production methods, continuous development and innovation, rigorous inspection processes, and strategic planning in navigating the competitive organic market. By integrating organic principles with sound business practices, these enterprises aim to achieve both economic viability and environmental stewardship, highlighting the sector's dynamic and evolving nature.

Topic 3: "Organic Production"

Represents a complex conceptual construct encompassing various aspects of organic agricultural practices and systems. The comprehensive analysis of the collected literature revealed a multifaceted approach to this topic, with numerous word groups converging under this thematic umbrella. The breadth of content covered within this topic is particularly noteworthy, as it spans multiple interconnected systems and processes integral to organic agriculture. These include:

1. Farming Systems: This refers to the holistic approach to agricultural production that adheres to organic principles, emphasizing ecological balance and biodiversity.

2. Food Systems: This encompasses the entire network of activities related to organic food production, distribution, and consumption, reflecting a broader perspective on the organic sector's impact.
3. Production Systems: This focuses on the specific methodologies and techniques employed in organic production, including crop rotation, natural pest management, and soil conservation practices.
4. Agricultural Systems: This term encapsulates the overarching framework of organic agriculture, including its social, economic, and environmental dimensions.
5. Organic Systems: This refers to the integrated approach to organic production, considering the interrelationships between various components of the organic farm ecosystem.
6. Inspection Systems: This highlights the crucial role of monitoring and verification processes in maintaining the integrity of organic production.
7. Organic Certification: This emphasizes the formal recognition and validation of organic production practices, essential for market access and consumer trust.
8. Sustainable Development: This concept underscores the alignment of organic production with broader sustainability goals, emphasizing long-term ecological balance and socio-economic viability.
The prevalence of these diverse yet interrelated concepts within Topic 3 underscores the holistic nature of organic production. It suggests that organic agriculture is not merely a set of production techniques but a comprehensive approach to food and farming that integrates environmental stewardship, social responsibility, and economic viability. This expansive conceptualization of organic production reflects the evolving understanding of organic agriculture as a multidimensional paradigm. It encompasses the practical aspects of cultivation and the regulatory frameworks, market dynamics, and broader societal implications associated with organic food and farming systems. The richness and diversity of content within this topic highlight the complexity of organic production and its far-reaching implications across various domains of agricultural science, environmental studies, and sustainable development.

Topic 4: "Organic Use"
Represents a conceptual framework elucidating the intricate relationships between organic utilization and methodological approaches in agriculture and related fields. This framework serves as a critical lens to examine the multifaceted aspects of organic and inorganic substance interactions, chemical residue management, and the application of organic compounds in various contexts.
The concept structure encompasses several key elements:
1. Inorganic and Organic Substances:
This component explores the interplay between inorganic and organic materials in agricultural systems. It examines how these substances interact, their roles in soil health, and their impacts on crop productivity and environmental sustainability.
2. Chemical Residue:
This aspect focuses on chemical residues' presence, persistence, and management in organic agricultural systems. It addresses the challenges of maintaining organic integrity while managing potential contamination from external sources.
3. Volatile Organic Compounds (VOCs):
The framework incorporates the study of VOCs in organic agriculture, including their natural occurrence in plants, their roles in plant defense mechanisms, and their potential applications in pest management strategies.
4. Organic Solvents:
This element examines the use and implications of organic solvents in various agricultural processes, including extraction methods for natural products and potential alternatives in organic farming practices.
5. Organic Acids:
The framework includes the study of organic acids, their roles in soil chemistry, plant nutrition, and potential applications in organic pest and disease management.
6. Compound Use:
This component investigates the strategic application of various organic compounds in agriculture, including their roles in soil amendment, crop protection, and post-harvest treatments.
7. Concentration Rate:
The concept structure addresses the importance of concentration rates in applying organic substances, exploring optimal levels for efficacy and safety in organic agricultural systems.
The "Organic Use" framework provides a comprehensive platform for understanding organic substances' complex interactions and applications in agriculture. It offers valuable insights that extend beyond traditional organic farming practices, potentially informing research and development in related fields such as agroecology, sustainable chemistry, and environmental science. This conceptual structure can guide future studies, facilitating interdisciplinary research that bridges the gap between organic

agriculture and other scientific domains. It encourages a holistic approach to understanding organic use, emphasizing the need to consider both the beneficial applications and potential challenges associated with organic substances in agricultural and environmental contexts. This framework can contribute to the development of more sustainable and efficient agricultural practices and inform policy decisions related to organic standards and regulations by providing a structured approach to examining organic use. Future research guided by this concept structure may lead to innovations in organic farming techniques, improved understanding of ecosystem dynamics, and the development of novel organic-based solutions for agricultural challenges.

3.4 DataVisualization

The interactive diagramming is a highly effective way to present the results of topic models. In this regard, the pyLDAVis package [13] was used to create 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 pyLDAvis visualization method is that users can adjust the relevance of words in a topic using a slider [18], [13]. 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[19].

The circles on the left panel show the overall view of the model which would allow users to easily understand the relationship between topics and their related strengths [20], [21], [22].

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.

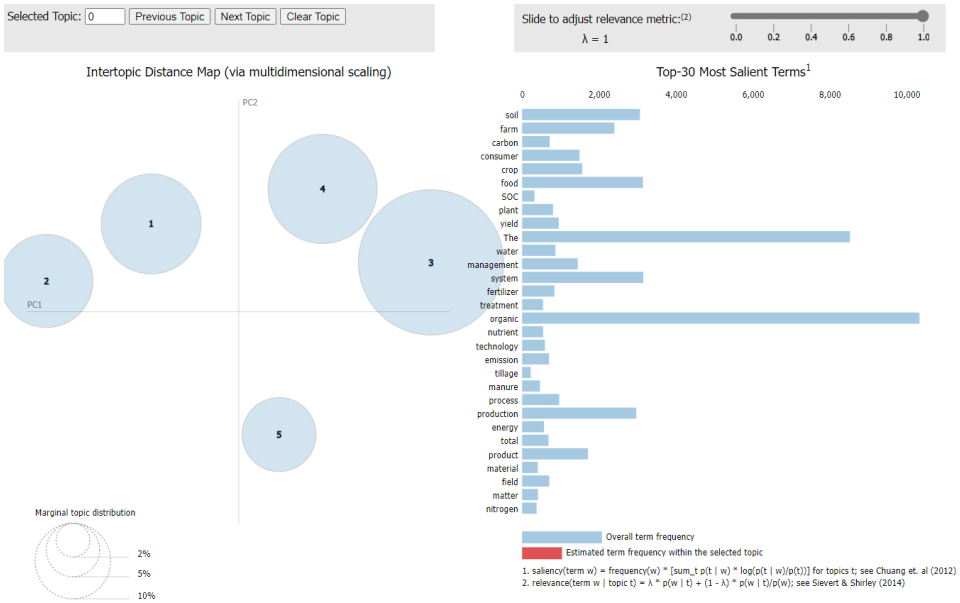

Figure14. Show pyDAVis topic model results.
https://ischool.kku.ac.th/bibliometric/lda_organic_agriculture.html

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 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 for 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 basically 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).

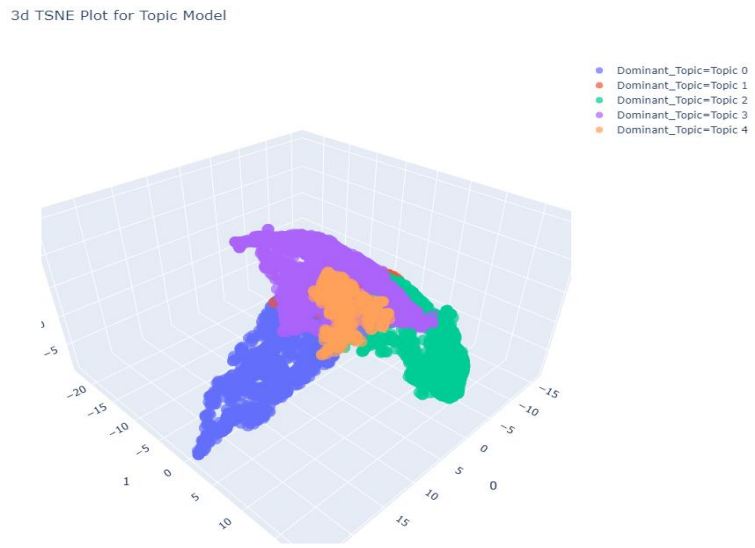

Figure 15. The Topic Distribution representing in 3-Dimensions

CONCLUSION

The analysis of research on organic agriculture and entrepreneurs from the barometric data through the text analysis technique with the PyCaret library in Python, 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.

In this study, the Bibliographic data were analyzed with text analysis and topic modelling techniques on abstracts from 4,327 collected articles from the Scopus database with analyzing 5 different aspects. The Python was employed to perform topic modelling and evaluate the performance of each topic. The model and present figure of the results clearly show the information which emphasizes 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 which were able to analyze and process accurately and precisely and can 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 limit for reading research results that might not cover the entire agricultural science. Therefore, future research should explore additional fields of science to get more comprehensive results.

ACKNOWLEDGEMENTS

This research was supported by the Graduate School of KhonKaen University, Thailand, under Grant No. 631S217-C. Additionally, this work received funding through a scholarship from the Research Assistant Program at KhonKaen University (Grant No. RA2566-12).

REFERENCES

[1] Office the Small and Medium Enterprises. (2020). "Report of Strategy and action plan for promoting small and medium enterprises in the agricultural industry." [Online]. Available: https://www.sme.go.th/upload/mod_download/download-20191022060242.pdf [Accessed: Nov. 20, 2023].

[2] Research Institute of Organic Agriculture. (2021). "Report of The World of Organic Agriculture 2021." [Online]. Available: http://www.pad.moi.go.th/images/formdownload/.pdf [Accessed: Nov. 20, 2023].

[3] National Organic Agriculture Development Committee, "National Organic Agriculture Development Strategy (2017-2021)," Bangkok: Office of Agricultural Economics, 2017.
[4] "Announcement on the National Economic and Social Development Plan No. 13," Royal Gazette, vol. 139, Special Section 258D, Oct. 24, 2022.
[5] Agricultural Research Development Agency. (2018). "Modern agriculture has transformed into a full-fledged businessman." [Online]. Available: https://www.arda.or.th/knowledge_detail.php?id=24 [Accessed: Nov. 20, 2023].
[6] M. Blake and S. Wijetilaka. (2015). "5 tips to grow your startup using SWOT analysis." [Online]. Available: http://www.afr.com/it-pro/5-tips-to-grow-your-startup-usingswot-analysis-20150226-13pkj5 [Accessed: Aug. 20, 2023].
[7] S. Joo et al., "Agriculture activity ontology: an ontology for core vocabulary of agriculture activity," in Proc. Int. Semantic Web Conf. (Posters & Demos), 2016, pp. 320-335.
[8] H. Jelodar et al., "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," Multimedia Tools and Applications, vol. 78, pp. 15169-15211, 2019.
[9] S. Wangsunthornchai, "Using the advanced search techniques of internet search engines," Journal of Library & Information Science, vol. 16, no. 1, pp. 21-36, 1998.
[10] M. Ali. (2020). "PyCaret - An open source low-code machine learning library." [Online]. Available: https://www.linkedin.com/pulse/pycaret-open-source-low-code-machine-learning-ali-mma-cpa-cma [Accessed: Aug. 20, 2023].
[11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," Journal of Machine Learning Research, vol. 3, pp. 993-1022, 2003.
[12] C. Sievert and K. Shirley, "LDAvis: A method for visualizing and interpreting topics," in Proc. Workshop on Interactive Language Learning, Visualization, and Interfaces, 2014, pp. 63-70.
[13] B. Grun and K. Hornik, "Topic models: an R package for fitting topic models," Journal of Statistical Software, vol. 40, no. 13, pp. 1-30, 2011.
[14] S. B. Cho, S. Shin, and D. S. Kang, "A study on the research trends on open innovation using topic modeling," Informatization Policy, vol. 25, no. 3, pp. 52-74, 2018.
[15] O. Koltsova and S. Koltcov, "Mapping the public agenda with topic modeling: the case of the Russian LiveJournal," Policy Internet, vol. 5, no. 2, pp. 207-227, 2013.
[16] D. Elgesem, L. Steskal, and N. Diakopoulos, "Structure and content of the discourse on climate change in the blogosphere: the big picture," Environ. Commun., vol. 9, no. 2, pp. 169-188, 2015.
[17] X. Li and L. Lei, "A bibliometric analysis of topic modelling studies (2000–2017)," Journal of Information Science, vol. 47, no. 2, pp. 161-175, 2021.
[18] J. Chuang, C. D. Manning, and J. Heer, "Termite: Visualization techniques for assessing textual topic models," in Proc. Int. Working Conf. Advanced Visual Interfaces, 2012, pp. 74-77.
[19] C. B. Asmussen and C. Møller, "Smart literature review: a practical topic modelling approach to exploratory literature review," J Big Data, vol. 6, no. 93, 2019.
[20] F. Diego, G. Mario, G. David, and M. L. Sergio, "Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach," IEEE Access, vol. 8, pp. 1-1, 2020.
[21] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," Journal of Business Research, vol. 133, pp. 285-296, 2021.
[22] T. Ta-Kham et al., "An Analysis of Specialty Coffee Research Publication Using a Bibliometric Method," Journal of Information Science, vol. 39, no. 1, pp. 2-19, 2021.