ISSN: 2508-7894
KJAI website: http://accesson.kr/kjai
doi: http://doi.org/10.24225/kjai.2024.12.3.17

A Study on the Development of LDA Algorithm-Based Financial Technology Roadmap Using Patent Data

Koopo KWON[1], Kyounghak LEE[2]

Received: August 05, 2024. Revised: August 19, 2024. Accepted: September 05, 2024.

Abstract

This study aims to derive a technology development roadmap in related fields by utilizing patent documents of financial technology. To this end, patent documents are extracted by dragging technical keywords from prior research and related reports on financial technology. By applying the TF-IDF (Term Frequency-Inverse Document Frequency) technique in the extracted patent document, which is a text mining technique, to the extracted patent documents, the Latent Dirichlet Allocation (LDA) algorithm was applied to identify the keywords and identify the topics of the core technologies of financial technology. Based on the proportion of topics by year, which is the result of LDA, promising technology fields and convergence fields were identified through trend analysis and similarity analysis between topics. A first-stage technology development roadmap for technology field development and a second-stage technology development roadmap for convergence were derived through network analysis about the technology data-based integrated management system of the high-dimensional payment system using RF and intelligent cards, as well as the security processing methodology for data information and network payment, which are identified financial technology fields. The proposed method can serve as a sufficient reason basis for developing financial technology R&D strategies and technology roadmaps.

Keywords : Patent Analysis, Financial Technology, Clustering Techniques, Trend Analysis, Network Analysis, Technology DevelopmentRroadmap, LDA

Major Classification Code: Artificial Intelligence, etc

1. Introduction

Since advanced technologies such as big data, artificial intelligence, and IoT developed, the overall procedure of the financial business process become automated, and various financial technologies are especially notable for the convenience and simplicity of users (Ahmed, 2020). Financial technologies used in various fields are being developed to minimize users' risk burden and cost reduction by securing transparency and stability. It is predicted that users will be guaranteed convenience through financial technology and companies will be able to build various business models (Pompella & Costantino, 2021).

Financial technology is a convergence technology field that includes various advanced technologies such as artificial intelligence, blockchain, big data, and cloud. The

* The present Research has been conducted by the Research Grant of Kwangwoon University in 2023.
1 First Author. Associate Professor, Department of Shipping and Air Cargo & Drone Logistics, Youngsan University, Republic of Korea, Email: koopo.kwon@ysu.ac.kr
2 Corresponding Author. Associate Professor, IACF, Kwangwoon University, Republic of Korea, Email: goldbug@kw.ac.kr

core of financial-related technology is that it provides service to users through data safely collection and analysis (Awotunde, 2021). Technologies in various fields must be systematically developed to provide stable financial services to users with a financial technology-based business model (Das, 2019). The rapid development of financial technology has not yet led to systematic development. Therefore, the financial technology development roadmap is crucially needed.

Convergence and combination between technologies are essential in the series of processes for developing financial technologies applied to various advanced technologies. Patent documents can be used to design a systematic roadmap based on information on technology. Since patent documents contain detailed descriptions and technical elements of the technology field, it is possible to analyze technology trends and establish a technology development roadmap by predicting future technologies for prior patent analysis. In addition, it is easy to predict the technology development process and new technology through patents applied in the past and currently applied (Park et al., 2005). Thus, this study intends to establish a future financial technology prediction and roadmap by analyzing technology development status using financial technology patents.

2. Theoretical Background

2.1. Current Status of Fintech Development

Financial technology began with the advent of credit cards, the computerization of bank transactions, and the development of the Internet with advanced digital technologies opened the fintech society. Financial technologies including advanced technologies such as blockchain, financial security, bio-authentication, artificial intelligence, and big data can greatly contribute to improving the accessibility and reliability of financial consumers. Based on this fact, to secure competitiveness in the financial industry, many companies are establishing technology development strategies for detailed financial technology fields.

Financial technology subfields can be categorized into payment systems, advisory services, and fund management. Utilizing financial technology enables the creation of diverse business values. Payment systems seek to establish a cashless transaction environment and various companies are developing related systems. These companies create proprietary payment applications and develop decentralized smart contract systems using blockchain technology. Robo-advisors, advisory services incorporating advanced technologies such as artificial intelligence, big data, and machine learning, provide effective services by offering personalized recommendations.

Financial technologies encompass a broad spectrum of technologies applied to financial processes (Goldstein et al., 2019). To develop various FinTech solutions, systematic development and integration of multiple sub-technologies are necessary. For instance, ensuring secure financial transactions both online and offline requires robust financial security measures to build reliability. Personalized recommendation systems need precise analytical techniques based on payment information. However, from the perspective of overall financial processes, utilizing transparent payment information in recommendation systems means that all processes must be systematically developed as a sequence of interconnected stages. Since they are interrelated processes, the isolated development of individual technologies is insufficient for securing competitiveness in the financial industry.

Therefore, to develop a technological roadmap for the financial industry, it is essential to establish the roadmap based on an accurate status investigation utilizing patent documents.

2.2. Patent Analysis

2.2.1. Technological Analysis through Latent Dirichlet Allocation (LDA)

The Latent Dirichlet Allocation (LDA) algorithm, a topic modeling method, is a methodology for finding various hidden topics within documents through clustering. Numerous research cases have used this algorithm to discover technological trends and to derive technology development strategies. In social network service, topics were segmented based on frequently occurring words to identify contemporary issues, and strategies for each field were developed (Lee et al., 2017). Examples include not only patent documents but also research papers. The LDA algorithm was applied to analyze trends and technological sectors in solar energy technology and identify changing technological fields. By analyzing core keywords based on these sectors, the study derived technology strategies for R&D (Yeo & Jeong, 2020). Likewise, many studies have confirmed the potential for diverse technological fields using the LDA technique, and this study aims to analyze financial technology applying the LDA technique.

2.2.2. Technological Analysis through Network Analysis

Network analysis is a methodology that visually represents nodes (actors) as technological elements and quantitatively analyzes these nodes using various metrics. It has been applied in the social sciences to analyze relationships between people or specific elements. For instance, network analysis of cited references in periodicals literature has been used to propose specific future plans (Ji & Qiu, 2002). In the field of research, network analysis was employed to analyze citation relationships between journals and authors, helping to identify research directions and the journals characteristics.

3. Research Methods and Materials

3.1. Unstructured Data Analysis

3.1.1. Data Preprocessing Procedure

The primary goal of data preprocessing is to remove unnecessary information and extract data that aligns with the objective. This process directly affects the performance and accuracy of the model and is crucial for the accuracy and reliability of subsequent data analysis. Figure 1 indicates the data preprocessing procedure. During data preprocessing, stemming and lemmatization techniques are used. Stemming is a method that analyzes words in simplest morphological form, while lemmatization analyzes words in their base dictionary form. Since lemmatization requires knowledge of the part of speech to yield accurate results, this study applies the more accurate stemming technique (Pradha et al., 2019). The refined words after the data preprocessing procedure cannot be immediately utilized due to the presence of unnecessary words. Accordingly, a metric to evaluate the significance of each word in each document is necessary.

This study applies Term Frequency-Inverse Document Frequency (TF-IDF) to ensure objectivity and measure the importance of each word. The TF-IDF technique calculates the weight of words in a document based on multiplying the TF value and the IDF value, thereby selecting important keywords (Aizawa, 2003).

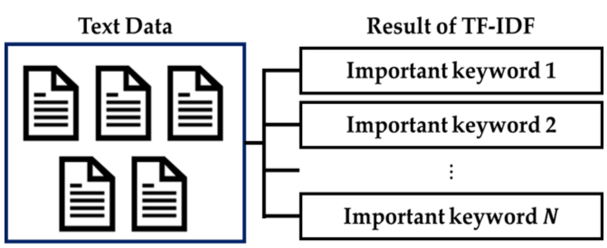


Figure 1: Data Preprocessing Procedure

3.1.2. LDA (Latent Dirichlet Allocation)

LDA (Latent Dirichlet Allocation) technique is a probabilistic topic modeling technique that identifies latent topics and themes within documents collected through natural language processing technology. Topic modeling is a method that selects candidate keywords from similar documents and evaluates their importance to extract significant and latent concepts which are not explicitly included in the documents (Ko, 2017). In this method, each topic clusters the documents with a combination of keywords from the original document content.

The LDA algorithm uses Bayesian statistical inference, typically employing the Dirichlet distribution. This inference method uses the prior probability of the target and additional provided information to infer the posterior probability, serving as a basis for extracting topics from pre-surveyed documents and keywords. The relational expression for the inference method is demonstrated in Equation 1, where the observed word is denoted by w_i, the latent topic by k, and the probability that a specific observed word reflects a specific latent topic is defined as β_{ik} (Blei & Ng, 2003).

$$Z_k = \sum_{i=1}^{n} \beta_{ik} \cdot w_i \quad (1)$$

The operational principle of the LDA technique is illustrated in Figure 2. LDA probabilistically derives the topics that constitute the documents and iteratively selects and eliminates the words included in the topics to generate new documents. The goal is to estimate all variables using the hidden parameter α and the latent parameter β. The final objective is to infer the probability distribution that defines the topics for a specific document and which represents the words of the topics.

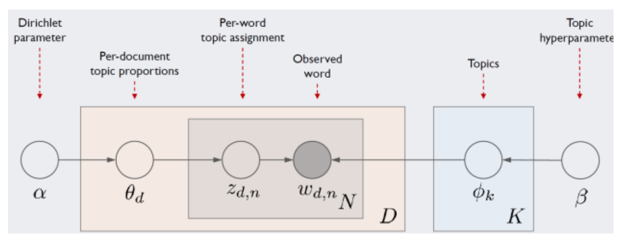


Figure 2: Operating Principle of the LDA Technique

The present study goes beyond merely classifying simple finance-related topics through the LDA technique. It can also extract keywords that have close connections to these topics. Ultimately, these keywords can be used to interpret financial technologies and define their values.

3.2. Patent Analysis

3.2.1. Trend Analysis

Trend analysis is applied to analyze technological prospects. It uses time-series data to develop mathematical models that explain the fundamental relationships of observed values and predict future trends (Zhang, 2003). By examining the direction and magnitude of the slope, future trends can be identified. Therefore, this study analyzes trends in various financial technology sectors based on the annual development proportions.

3.2.2. Network Analysis

Network analysis structurally represents entities as nodes and their relationships as links, ranging from individuals to groups. This analysis quantitatively examines the confidence from a specific entity to others within the network structure and further the evolution process. Various network centrality metrics are used to analyze the influence of nodes. The representative network centralities are degree centrality, closeness centrality, and betweenness centrality, which help to identify important nodes in the network and determine whether they are structurally central (Yan & Ding, 2009).

4. Results

4.1. Patent Data Collection and Data Preprocessing

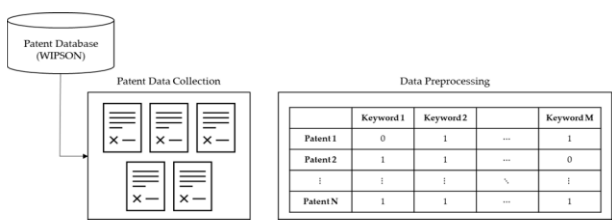


Figure 3: Data collection and Preprocessing Procedure

This study utilized the Wipson database to collect financial technology patents. Extracting published/registered patents in the United States, Europe, China, and Korea from 1974 to 2021. Search queries were formulated based on financial technology-related reports and previous studies. A total of 3,246 patents were extracted, and through a selection process involving financial technology experts, 1,558 valid patents were analyzed. The preprocessing procedure for unstructured data is demonstrated in Figure 3. During the stemming process, trimming the meanings of words, the Stemming technique was applied. Using the TF-IDF algorithm, the top 200 words were selected as key terms to construct a patent-keyword matrix, as demonstrated in Table 1.

Table 1: Patent-Keyword Matrix (1,558 X 200)

	payment	user	mobil	transact	···	wireless
Patent 1	1	1	1	0	···	0
Patent 2	0	1	1	0	···	0
Patent 3	1	1	0	0	···	1
⋮	⋮	⋮	⋮	⋮	⋱	⋮
Patent 1556	0	0	0	0	···	1
Patent 1557	0	0	1	1	···	1
Patent 1558	1	0	1	1	···	1

4.2. Clustering Financial Technology and Analysis of technology status

4.2.1. Technology Clustering Using LDA Technique

In this stage of the study, the LDA technique is applied to the patent-keyword matrix to cluster financial technology. The most crucial parameter in the LDA analysis is the number of topics, K, which is typically determined as quantitative metrics such as word complexity and the silhouette method. However, if the number of topics is too high during the technology evaluation process, the analysis may become too dispersed, reducing the significance of each topic (Zhang et al., 2021). As a result, based on the experience of financial experts, this study sets the number of LDA topics to 10 for the analysis. The ten topics derived from the LDA analysis and the important keywords originating from the word frequency in each topic are summarized in Table 2.

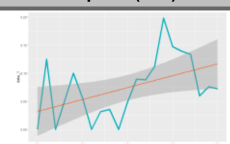
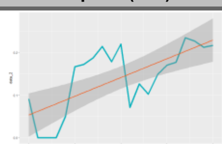
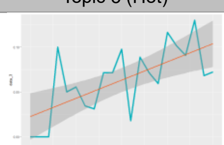
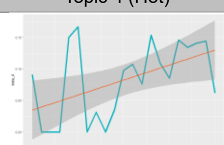
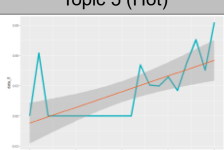
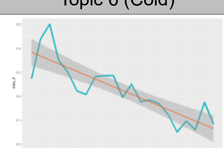
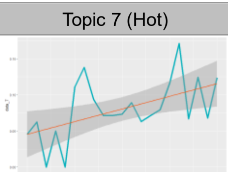
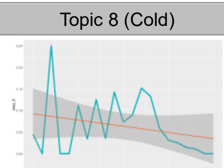
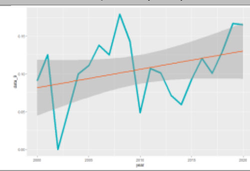
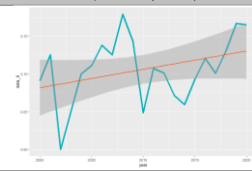
Table 2: LDA Analysis Results (K=10)

Topic	Definition and Core Keywords					
Topic 1	Online and Internet-Based Payment Systems					
	Payment	termin	mobil	code	user	secur
Topic 2	Data Information and Network Payment Security Processing Methods					
	manag	onlin	oper	transact	time	sale
Topic 3	Code Scanning-Based Payment Technology					
	payment	scan	vend	selfservic	mobil	purchas
Topic 4	Bluetooth-Based Payment Systems					
	user	mobil	termin	electr	payment	bluetooth
Topic 5	Multi-User Sharing-Based Financial Management and Supervision Systems					
	manag	onlin	communic	storag	interact	wireless
Topic 6	RF-Based Wireless Communication Payment Systems					
	card	account	purchas	payment	transact	credit
Topic 7	NFC-Based Online and Payment Technology					
	payment	transact	user	certif	transmit	idenrifi
Topic 8	Advanced Payment Systems Using RF and Smart Cards					
	phone	payment	communic	bank	nfc	wireless
Topic 9	Electronic Payment System Applications					
	payment	purchas	advertis	time	onlin	search
Topic 10	Online and Mobile Transfer Payment Systems					
	communic	servic	program	oper	connect	transmit

4.2.2. Analysis of Technological Development Status by Topic

To assess the technological status of the 10 topics derived from the LDA analysis, the trend of the annual proportion of each topic was analyzed based on the LDA results. This analysis method, which is one of the Technology and Innovation Management (TIM) methods, identifies the development status of each topic. The development status is categorized into rising trends (Hot topics) and declining trends (Cold topics) (Lee & Kang, 2018). This study applied the TIM method to patent data post-2000 to analyze the technological development status of each topic. The results are summarized in Table 3. Linear regression was applied to determine the slope values to distinguish Hot topics and Cold topics. Seven topics were identified as Hot topics, while three were identified as Cold topics. Among the Hot topics, Topic 2, which had the highest slope, was found to be the most active field, indicating that the development of a technology roadmap for this area is the most urgent.

Table 3: Analysis of Technological Development Status by Topic

	Topic 1 (Hot)	**Topic 2 (Hot)**
Graph		
Trand result	Slope = 0.004242	Slope = 0.008793
	Topic 3 (Hot)	Topic 4 (Hot)
Graph		
Trand result	Slope = 0.004054	Slope = 0.004751
	Topic 5 (Hot)	Topic 6 (Cold)
Graph		
Trand result	Slope = 0.0031245	Slope = -0.015926
	Topic 7 (Hot)	Topic 8 (Cold)
Graph		
Trand result	Slope = 0.003504	Slope = -0.002876
	Topic 9 (Hot)	Topic 10 (Cold)
Graph		
Trand result	Slope = 0.002432	Slope = -0.12097

4.2.3. Analysis of Technological Similarity by Topic

The technological similarity analysis for each topic derived from the LDA analysis assesses the potential for future technology convergence. The cosine similarity of each topic was measured using the patent-keyword matrix to verify technological similarity, and Table 4 demonstrates the result. The analysis of similarity between topics reveals that Topic 5 and Topic 8 are highly similar to most other topics. Specifically, the similarity between Topic 5 and Topic 8 is the highest, judged to be a high potential for technology convergence.

Table 4: Top 5 Similarities by Topic

Rank	Topic A	Topic B	Cosine Similarity
1	Topic 5	Topic 8	0.7058511
2	Topic 2	Topic 8	0.5549681
3	Topic 1	Topic 5	0.5115268
4	Topic 5	Topic 7	0.5092158
5	Topic 5	Topic 10	0.5065880

4.3. Clustering and Analysis of Financial Technology

Based on the topic extraction and inter-topic similarity analysis results obtained using LDA, Topic 2 expects to be the most prominent topic. In contrast Topic 8 exhibits the highest similarity. Whereas the future trend for Topic 8 appears to be declining, it is anticipated to merge with topics such as Topic 2 and Topic 5. Building on these results, the key technological elements of Topic 2 and Topic 8 are identified using a keyword network, and a technology roadmap through technological convergence is developed.

4.3.1. Keyword Network Analysis by Topic

Figure 4 shows the network analysis results for Topic 2. Through this network analysis, it was identified that "mobil," "user," and "communic" were identified as key keywords with high betweenness centrality. This implies that payment and security among users connected via mobile are essential technological elements in the field of data information and network payment security methodology. Therefore, the development of technologies including these elements is necessary for the advancement of Topic 2. Related technological elements include "messag," "network," "oper," "card," and "secur," highlighting the necessity of developing auxiliary technologies to manage and protect networks among various users.

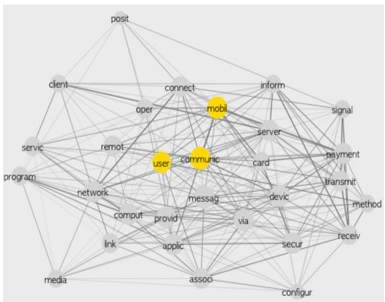


Figure 4: Network Analysis Results of Topic 2

Table 5: Network Analysis of Topic 2 (Top 5)

keyword	betweenness	centrality	closeness
servic	24.279237	54	0.9655172
internet	22.590284	50	0.9310345
onlin	20.176835	48	0.9137931
market	18.150687	48	0.9137931
time	17.022432	46	0.8965517

Figure 5 shows the network analysis results for Topic 8. Through this network analysis, “onlin”, “servic”, and “internet” were identified as key keywords with high betweenness centrality. It indicates that in high-dimensional payment systems that use RF and intelligent cards, the online transmission of information and payment systems are essential technological elements. Therefore, the development of technologies, including these elements, is necessary for the advancement of Topic 8. Related technological elements include “time”, “inform”, “market”, “consum”, and “price”, highlighting the necessity of developing auxiliary technologies for real-time information transmission of store sales data.

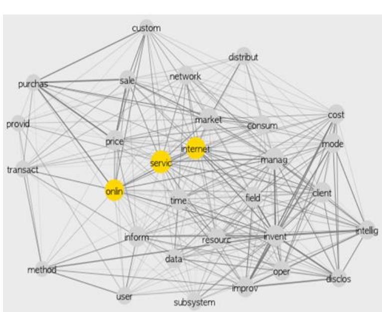


Figure 5: Network Analysis Results of Topic 8

Table 6: Network Analysis of Topic 8 (Top 5)

keyword	betweenness	centrality	closeness
user	23.131358	38	0.8275862
communic	22.180902	44	0.8793103
mobil	21.962138	42	0.862069
messag	20.393978	44	0.8793103
comput	19.062981	38	0.8275862

4.3.2. Keyword Network Analysis Between Topics

The keyword network analysis aims to identify technologies to integrate Topic 2 and Topic 8. For investigation, a network was constructed based on the patent-keyword matrices corresponding to Topics 2 and 8, and central technologies were extracted using the betweenness centrality metric. Figure 6 shows the results of the fusion network analysis for Topics 2 and 8. Through this network analysis, “servic”, “network”, “data”, “communic”, “payment”, and “user” were identified as key technologies to have high betweenness centrality.

This implies that to integrate the technologies of Topic 2, which involves security processing methodologies for data information and network payments, and Topic 8, which involves high-dimensional payment systems using RF and intelligent cards. Related technological elements including "price," "internet," "consum," "resourc," "oper," "connect," "inform," "mobil," "transact," and "link," highlight the necessity of developing connected technologies for accessing and managing transaction information between consumers and sellers through various devices such as the internet and mobile phones.

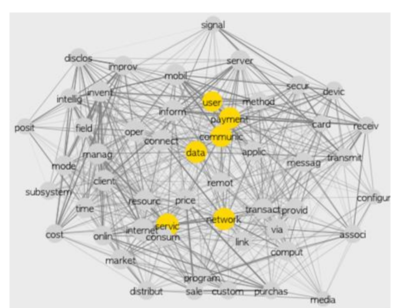


Figure 6: Fusion Network Analysis of Topic 2, 8 (Top 5)

Depending on the results, the financial technology development roadmap comprises two stages. Table 5 summarizes results of the analysis, used to establish the technology development roadmap. Stage 1 of the technology roadmap involves the development of “Security Processing Methodologies for Data Information and Network Payments” and “High-dimensional Payment Systems Technology using RF and Intelligent Cards”. The roadmap should include technologies such as mobil, user, communic, message, network, oper, card, and secure to address payment and security issues among various users to develop “Security Processing Methodologies for Data Information and Network Payments”. Additionally, the roadmap should present technologies related to online, service, internet, and further utilize information related to time, inform, market, consumer, and price to develop the “High-dimensional Payment Systems Technology using RF and Intelligent Cards”.

Consequently, these two technologies need to be integrated to advance financial technology. For this achievement, the fusion technology should develop “Communication and Payment Systems between Users (Consumers, Sellers) and Service Technologies Utilizing Sales Data” and propose a technology roadmap that applies technologies for accessing and managing various devices utilizing these technologies.

.

Table 7: Summary of Network Analysis Results

Topic	Key Technological Elements	Related (Auxiliary) Technological Elements
Topic 2	mobil, user, communic	messag, network, oper, card, secur
Topic 8	onlin, servic, internet	time, inform, market, consum, price
Topic 2, 8	servic, network, data, communic, payment, user	price, internet, consum, resourc, oper, connect, inform, mobil, transact, link

5. Conclusions

This study proposes a systematic financial technology development roadmap focusing on future-oriented financial technologies through patent analysis. Using stop-word processing and TF-IDF, text preprocessing was performed on key keywords, which were then subdivided into 10 topics using LDA topic modeling. These topics were analyzed to determine their current positions and trends in technology development. The results indicated that "Security Processing Methodologies for Data Information and Network Payments" has significant growth potential. We conducted an inter-topic similarity analysis to facilitate technology convergence, which revealed that "High-dimensional Payment Systems using RF and Intelligent Cards" could be integrated with the topic above.

The financial technology development roadmap was structured in two stages. The stage 1 technology development roadmap focuses on identifying the core technologies of each topic, highlighting that payment and security among mobile-connected users, online information transmission and payment systems are critical technological elements. The stage 2 technology development roadmap aims to integrate the two topics, concluding that service technologies based on communication and payment among users and sales data must be developed to establish the technology development roadmap.

The results of this study not only align closely with financial technology trends but also identify elements at the technological component level, making them applicable for R&D and national policy roadmap development. This study confirms that financial technology forecasting is feasible through patent analysis and denotes those identifying relationships among technological elements allows for deriving more specific technology development roadmaps rather than general conceptual roadmaps. Given that this research methodology can be applied to fields other than financial technology, it is expected to be useful for developing technology roadmaps for various other domains.

References

Ahmed, A. S., Kumar, M., & Ali, M. A. M. (2020). *Adoption of FinTech and Future Perspective: An Empirical Evidence from Bahrain on Digital Wallets.* In 2020 International Conference on Decision Aid Sciences and Application (DASA) (pp. 751-755).

Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. *Information Processing & Management, 39*(1), 45-65.

Awotunde, J. B., et al. (2021). *Fintech with Artificial Intelligence, Big Data, and Blockchain.* In Springer (pp. 107-139).

Blei, D. M., & Ng, A. Y. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993-1022.

Das, S. (2019). Opportunities and challenges of FinTech. Keynote Address Delivered at NITI Aayog's Fin-Tech Conclave*, New Delhi, 25*(3).

Goldstein, I., Jiang, W., & Karolyi, G. A. (2019). *To Fintech and Beyond. The Review of Financial Studies, 32(5), 1647-1661.*

Ji, J. L., & Qiu, E. (2002). Study on Quotation and Citation Form of Network Document [J]. *Information Science, 1.*

Ko, M.-S. (2017). Unstructured Data Processing Using Keyword-Based Topic-Oriented Analysis. *Journal of the Korea Information Processing Society, 6(11), 521-526.*

Lee, H., & Kang, P. (2018). Identifying core topics in technology and innovation management studies: A topic model approach. *Journal of Technology Transfer, 43*(5), 1291-1317.

Lee, I. S., Jang, J. H., & Yoo, K. H. (2017). *Twitter Data Analysis System using LDA model.* Proceedings of the Korea Information Processing Society Conference, *24*(2), 389-390.

Park, Y., Yoon, B., & Lee, S. (2005). The idiosyncrasy and dynamism of technological innovation across industries: Patent citation analysis. *Technology in Society, 27(*4), 471-485.

Pompella, M., & Costantino, L. (2021). *Fintech and Blockchain Based Innovation: Technology Driven Business Models and Disruption.* In Springer (pp. 403-430).

Pradha, S., Halgamuge, M. N., & Tran Quoc Vinh, N. (2019). *Effective text data preprocessing technique for sentiment analysis in social media data.* In 2019 11th International

Conference on Knowledge and Systems Engineering (KSE) (pp. 1-8). Da Nang, Vietnam.

Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science & Technology, 60*(10), 2107-2118.

Yeo, J.-S., & Jeong, Y. (2020). Pathway toward market entry of perovskite solar cells: A detailed study on the research trends and collaboration networks through bibliometrics. *Energy Reports, 6*(2), 2075-2085.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50,* 159-175.

Zhang, H., Daim, T., & Zhang, Y. P. (2021). Integrating patent analysis into technology roadmapping: A latent Dirichlet allocation based technology assessment and roadmapping in the field of Blockchain. *Technological Forecasting & Social Change, 167*, 1-13.