



XXXI International Scientific Conference
Strategic Management
 and Decision Support Systems
 in Strategic Management
Sm2026

Subotica (Serbia), May 18, 2026

Rade Božić

Faculty of Business Economics Bijeljina,
 University of East Sarajevo Bijeljina
 Bijeljina, Bosnia and Herzegovina

rade.bozic@fpe.ues.rs.ba

Mirjana Marić

Faculty of Economics in Subotica, University of
 Novi Sad
 Subotica, Serbia

mirjana.maric@ef.uns.ac.rs

Participation (direct/virtual):

JEL:

Olivera Grljević

Faculty of Economics in Subotica, University of
 Novi Sad
 Subotica, Serbia

olivera.grljevic@ef.uns.ac.rs

Nebojša Taušan

Faculty of Economics in Subotica, University of
 Novi Sad
 Subotica, Serbia

nebojsa.tausan@ef.uns.ac.rs

direct

C38, Z30

Željko Vukićević

Faculty of Economics, University of Kragujevac
 Kragujevac, Serbia

zeljko.vukicevic@ef.kg.ac.rs

HOSPITALITY AND TOURISM RESEARCH DURING GLOBAL INSTABILITY: A BERTOPIC ANALYSIS

Abstract: Hospitality and tourism represent the sector most affected by global disruptions caused by pandemics, energy crises, or political disturbances. In this paper authors examine how instability shapes research priorities and responses in hospitality and tourism research between 2019 and 2023, a period characterized by global pandemics. Topic modelling is utilized as the main method applied on abstracts of 1,100 scientific publications from this domain. These publications are collected from Web of Science and Scopus citation databases. Topic modelling, a natural language processing technique, uncovers hidden thematic structures in large text collections and corpora. This study utilized BERTopic to uncover research topics studied within an observed time interval. BERTopic uncovered 23 dominant topics in hospitality and tourism research, which emerged during the pre-crisis, crisis, and the adaptation period. Most prominent topics discovered include customer experience and review, energy systems modelling, consumption, brand loyalty and satisfaction and destination image and behavior.

Keywords: topic modeling, BERTopic, NLP, hospitality, tourism

INTRODUCTION

Global disruptions can take various forms and influence numerous sectors and multiple aspects of social and economic life. The hospitality and tourism sector is highly vulnerable to different types of crises, including pandemics, terrorism, political instability, natural disasters, and economic recessions (Kocak, Okumus & Altin, 2023; Leta & Chan, 2021; Liu, Pahrudin, Tsai & Hao, 2024; Sampaio, Farinha, Sebastião & Fernandes, 2023). Any factor that restricts human mobility has an impact on this sector and its overall performance and stability. The growth and development of many national economies depend directly on tourism and hospitality, and in some cases, it even represents the primary economic activity that generates value (Wijesekara et al., 2022). For example, COVID-19 outbreak led to 2.86 trillion USD losses, which represents more than 50% of total income in this area (Abbas, Mubeen, Iorember, Raza & Mamirkulova, 2021). Therefore, tourism represents an important research area that attracts considerable attention from the scientific community, and any change in its functioning represents a research challenge that becomes a relevant topic in academic studies (Leta & Chan, 2021; Liu et al., 2024).

Since the early 2000s, researchers have actively examined the vulnerability of the tourism and hospitality sector to various destabilizing events. During this period, attention was devoted to events such as terrorist attacks, bombings, tsunamis, and the SARS outbreak. Around the 2010s, research focus was shifted toward the impact of global financial crises on tourism demand and stability (Liu et al., 2024). More recently, the COVID-19 pandemic has significantly increased academic interest in crisis-related disruptions and highlighted the sensitivity of tourism and hospitality to global shocks (Liu et al., 2024; Wong, Kim, Liu & Grace Baah, 2025; Zhong, Sun, Law & Li, 2021). To better understand the shifting of research focus during the crisis period, this study focuses on topics which were examined in the tourism and hospitality literature during the COVID-19 pandemic. One possible approach to answering this question is the application of topic modeling (TM), a natural language processing (NLP) technique on published studies. Topic modeling is a statistical, NLP tool for extracting hidden semantic structures in large collections of documents (Blei, Ng & Jordan, 2003). It can also be described as a generative model in which documents are represented as mixtures of topics, while each topic is represented as a probability distribution over words (Alghamdi & Alfalqi, 2015). This modeling approach has been applied across various fields such as medical sciences, software engineering, geography, political science, marketing, sociology and digital humanities (Dieng, Ruiz & Blei, 2020; Jelodar et al., 2019). The most used topic modeling methods include Latent Dirichlet Allocation (LDA), Non-negative Matrix Factorization (NMF), Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (pLSA), BERTopic, Correlated Topic Models (CTM), Top2Vec, LDA2Vec (Hankar, Kasri & Beni-Hssane, 2025).

Compared with other models such as LDA, BERTopic shows better performance in terms of semantic relevance and contextual coherence (Sy, Maceda, Flores & Abisado, 2024). This method generates topics by converting documents into embedding representations, then reducing their dimensionality to optimize the clustering process, and finally extracting topics using a class-based variation of TF-IDF (Grootendorst, 2022). For this reason, the BERTopic is chosen as a topic modelling method in this study. Consequently, the aim of this paper is to apply BERTopic to identify topics that reflect the research focus and priorities in the field of hospitality and tourism during the period of global instability caused by the COVID-19 pandemic. The extracted topics are used to answer two research questions. These research questions are:

RQ1: Which topics dominate research in the field of hospitality and tourism between 2019 and 2023?

RQ2: Does the research focus shift on resilience and risk management after 2020?

The dataset used for answering these RQs consists of 1,212 research papers retrieved from the Scopus and Web of Science citation databases, published between 2019 and 2023, while after pre-processing of publications 1,100 remained for analysis.

The remainder of this paper is organized as follows. The first section presents the theoretical foundations of topic modeling, with a particular focus on the BERTopic method. *Materials and Methods* section presents data used in the study and the methodology that was applied. Third section presents the results of the modelling, followed by discussion section where focus is to present obtained findings from topics through time transition of COVID pandemic. Finally, the conclusion is presented.

1. TOPIC MODELING

Topic modelling, a natural language processing technique, reveals thematic structures from collection of texts that may not have been anticipated. It relies on unsupervised learning algorithms that detect patterns within textual data (Abdelrazek, Eid, Gawish, Medhat, & Hassan, 2023; Alghamdi & Alfalqi, 2015). By grouping texts according to semantic similarity and context, topic modelling forms clusters or topics (Kherwa & Bansal, 2020). BERTopic represents one of the emerging topic modeling methods in applied research in hospitality domain (Egger & Yu, 2022). Due to its advantages, such as it represents multilingual model for 50 different languages (Egger & Yu, 2022) or that it enables researchers to examine and observe trends and changes addressed in the texts (Rejeb, Rejeb, Simske, & Süle, 2025), it is selected for the case study presented in this paper.

The BERTopic framework implies three key steps (Rejeb et al., 2025; Grootendorst, 2022). After collecting the documents, the first step refers to vector representation of documents. For this purpose, BERTopic utilizes BERT embeddings and most often a pre-trained Sentence Transformer (SBERT) that embeds sentences in a high-dimensional vector space. The second step is the reduction of the dimensionality of the vector for more effective clustering. It is recommended to use the UMAP (*Uniform Manifold Approximation and Projection*) technique in order to preserve the revealed information and word meanings. The last step involves grouping the reduced embeddings using clustering algorithms, such as HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*), and generating topics per cluster using the c-TF-IDF technique (Grootendorst, 2022). The HDBSCAN technique in the model enables the storage of identified outliers in a separate cluster, avoiding forced assignment to dominant and meaningful clusters, which would damage the quality of the output (Angelov, 2020). Thus, compared to traditional

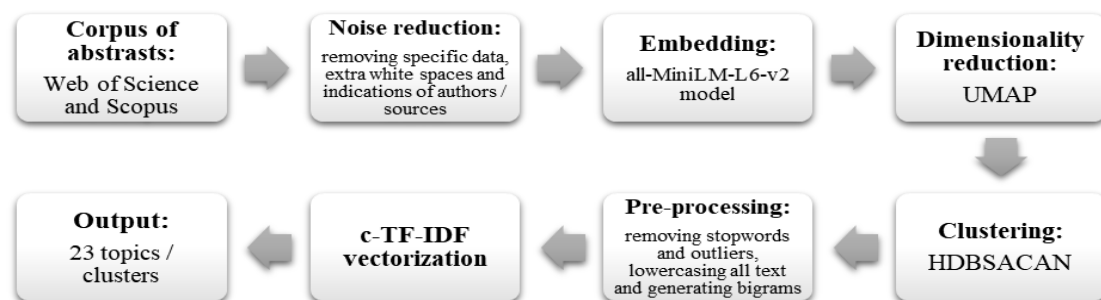
techniques, clustering is not performed on a centroid-based but on a class-based approach of calculated frequencies, the c-TF-IDF.

Although the mentioned steps are crucial for the application of the BERTopic model, the framework is extendable with additional steps, such as data pre-processing (Grootendorst, 2022; Rejeb et al., 2025), topics representation (Ma et al., 2025), or visualization and inspection of individual topics (Egger & Yu, 2022). Also, BERTopic model is not strictly restricted to the application of HDBSCAN and UMAP techniques. Other techniques and models can be used, depending on the need, goal, or application area of topic modelling. The flexibility and modularity of the model represents advantage over other models (Alammar & Grootendorst, 2024).

2. MATERIALS AND METHODS

Data is collected by querying Web of Science and Scopus, two citation databases complying with the highest publishing standards and cross-linking publications in numerous other citation databases. The study focuses on pandemics response of hospitality researchers examining customer/traveler experience, preferences, and perceptions by utilizing topic modeling as a technique. The search is set accordingly, using the combination of the following keywords and their alternative spellings with Boolean operators: “topic modeling” OR “topic modelling”, “tourism” OR “destination marketing” OR “tourist perceptions” OR “hospitality” OR “attractions” OR “tourists’ preferences” OR “customer satisfaction” OR “destination management”. The timeframe is restricted to period from 2019 to 2023, covering pre-pandemics, pandemics, and post-pandemics period, and only peer-reviewed journal articles written in the English language were included in the analysis.

The search resulted with 1,212 records. For the modelling purposes, we recorded paper title, author names, publication year, source title, and abstract in an Excel file. Two records were eliminated due to the lack of abstract and 110 records were duplicate records and removed, leaving a total of 1,100 records for further analysis. Prior embedding, as presented in Picture 1, the noise is eliminated by removing data specific to structured abstracts, such as “*Purpose:*”, “*Design/Methodology/Approach:*”, “*Research limitations/implications:*”, etc. Using regular expressions in Python, extra white spaces in the abstracts are removed, and all indications of authors or publications source. Such pre-processing was conducted to ensure the plain abstract text will be used as an input for topic modelling.



Picture 1: Topic modelling flow
Source: Authors

To numerically represent abstracts and capture their meaning, we utilized sentence transformers model *all-MiniLM-L6-v2*, which is optimized for semantic similarity tasks and clustering, an important criteria for selection of embedding model for topic modelling (Alammar & Grootendorst, 2024). Resulting embedding model is used for BERTopic modelling along with class based variant of TF-IDF for vectorization (Grootendorst, 2022), UMAP for dimensionality reduction (McInnes, Healy, Saul, & Großberger, 2018), and HDBSCAN for clustering (McInnes, Healy, & Astels, 2017). Following (Grootendorst, 2022) suggestion, a creator of BERTopic, light pre-processing is done after embeddings and prior processing topic representations: removal of stopwords, lowercasing all text, generating bigrams, and removing extremely rare words (that appear in fewer than 3 documents). Extracted semantic topics spanning five-year research in hospitality and tourism are used to examine the research response on global instability caused by pandemics.

3. RESULTS

3.1. Corpus statistics

The corpus comprises 1,100 abstracts collected from 594 different journals. Table 1 illustrates distribution of abstracts per year. Data indicates gradual shift in the number of papers, indicating the rise in relevance of topic modeling as analytical tool in hospitality and tourism sector.

Table 1: Distribution of abstracts per year

Year	Papers	Percentage	Cumulative %
2019	158	14.36	14.36
2020	185	16.82	31.18
2021	229	20.82	52.00
2022	286	26.00	78.00
2023	242	22.00	100.00

Source: Authors

Kern, et al. (2016) found that documents shorter than 500 tokens represent short text and, as illustrated in Table 2, corpus predominantly comprises short abstracts. On average abstract length is 206.83 tokens, i.e. words, while median is 195. This indicates right-skewed distribution and presence of atypically long abstracts. Percentile distribution confirms this – 75% of documents contain 237 tokens or less.

Table 2: Abstract characteristics

Abstract length statistics	Value
Mean	206.83
Std	95.43
Min	16
25%	151
50%	195.00
75%	237
Max	1,474

Source: Authors

Total number of tokens in the corpus is 229,522, while 11,625 are unique. Type-token ratio, calculated as the share of unique words in the total number of tokens, is 0.0506, which indicates very low lexical diversity, i.e. highly repetitive texts and vocabulary restriction. Only ~5% of total tokens used are unique.

3.2. BERTopic model

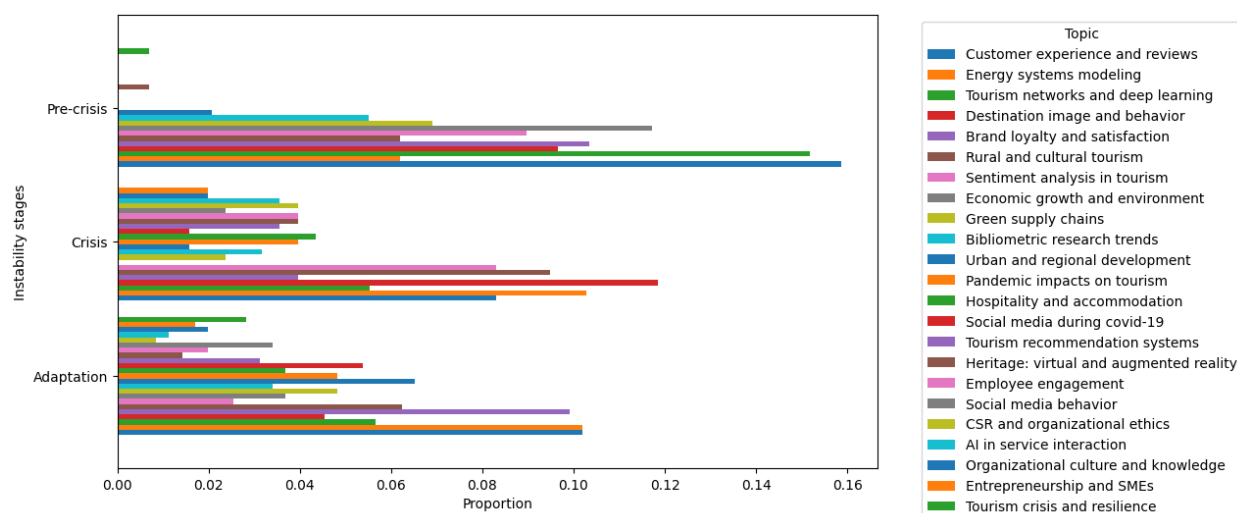
BERTopic modelling resulted in 23 topics presented in Picture 2 with top 10 keywords extracted during modelling based on their distribution over each topic (Grootendorst, 2022). Topic names are defined based on keyword analysis.



Picture 2: Top keywords per topic
Source: Authors

3.3 Topics analysis and interpretation

The instability period can be subdivided into stages. In this study, the year of publication was used to subdivide the publications into three stages and these stages are: pre-crisis, crisis, and adaptation. Pre-crisis refers to 2019, crisis period refers to 2020 and 2021, while adaptation period refers to 2022 and 2023. Picture 2 illustrates topic distribution by these three periods. The values are normalized to indicate the proportion of topic in each instability stage. Normalization allows comparison of the relative importance of topics within each period, independent of the number of papers.



Picture 3: Topic distribution by instability stages
Source: Authors

The distribution of topics indicates differences in research topics within the observed periods. In the period before the crisis caused by Covid-19, the research focus was on development topics. Researchers explored user experience based on the analysis of user online reviews (topic: Customer experience and reviews), ways to optimize tourism systems using deep learning (topic: Tourism networks and deep learning), as well as economic and ecological aspects of sector development (topic: Economic growth and environment).

In the period marked by the peak of crisis, topics characterized by the crisis emerged. These topics were not addressed in the previous period. The researchers are engaged in examining the public's perception of the risks associated with the pandemic and the impact on tourism (topic: Pandemic impacts on tourism), which is closely linked with studying social media communication during Covid-19 (topic: Social media during Covid-19) and sentiment analysis of these posts (topic: Sentiment analysis in tourism). In this stage of instability, the most pronounced topic that researchers explored relates to the image of tourist destinations and the behavior of tourists (topic: Destination image and behavior). Compared to other periods when employee engagement and dynamics were not studied or were studied on a significantly smaller scale, in the crisis phase this topic represents one of the emerging research trends (topic: Employee engagement), as well as corporate social responsibility and ethics in the organization (topic: CSR and organizational ethics).

In adaptation stage, topic addressing implementation of artificial intelligence (AI) in human-service interactions emerges (topic: AI in service interaction), as well as more pronounced research focus on crisis management and tourism resilience (topic: Tourism crisis and resilience). These topics reflect response of the scientific community to the emergence of the pandemic, during which the application and implementation of AI systems took momentum in addressing the challenges of the pandemic and to reduce interpersonal contacts. Also, a significant amount of research was done into examining directions to strengthen the sector's resistance to future instabilities. Such topics point to a pronounced digital transformation of the tourism and hospitality sector, which was fueled by global instability.

During the adaptation period and the crisis period, a topic dedicated to the study of personalized recommendation systems and their application in the tourism and hospitality sector appeared (topic: Tourism recommendation systems), which is also a response to the changed conditions. These two periods are also characterized by the emergence of studying possibilities for recovery through innovation and entrepreneurship (topic: Entrepreneurship and SMEs).

CONCLUSION

The COVID-19 outbreak acted as a global disruption, destabilizing numerous sectors and activities and particularly affecting the hospitality and tourism industry, which are highly vulnerable to such conditions. This situation has attracted significant attention in the scientific community, resulting in a substantial number of studies published between 2019 and 2023. This study presents an analysis of these documents to answer two research questions: which topics were most prominent in the studies, and how the focus of these topics evolved over the observed period. To address the research questions and identify the key topics discussed in the papers, a topic modeling technique was applied, specifically the BERTopic method.

A total of 1,210 papers from the Web of Science and Scopus databases were collected and pre-processed. After preparing the corpus, BERTopic was applied to generate 23 distinct topics. The results of topic modeling within the three observed periods indicate clear changes in research interests or focuses. In the pre-crisis period, researchers focused on user experience and tourism development, while in the crisis period there is a shift towards topics closely related to the pandemic and risk perception during the crisis, as well as ways to strengthen the sector's resilience. In the adaptation phase, the research focus shifts to the digital transformation of the sector, the application of artificial intelligence and innovations in tourism and hospitality. By analyzing research topics such as customer experience, crisis risk perception, and digital innovation, this study illustrates how topic modelling can assist managers in times of instabilities with concrete and actionable insights, such as directions for services adaptation to tourist preferences, improvement of safety and crisis planning, and implementation of technological innovations that respond to dynamic preferences.

Future research could expand the analysis by including the results sections of the analyzed publications, thus further refine the identified topics. Furthermore, applying alternative topic modeling methods and expanding the corpus to include papers from additional scientific databases could provide a more comprehensive basis for comparison and deeper interpretation of the results.

REFERENCES

- Abbas, J., Mubeen, R., Iorember, P. T., Raza, S., & Mamirkulova, G. (2021). Exploring the impact of COVID-19 on tourism: transformational potential and implications for a sustainable recovery of the travel and leisure industry. *Current Research in Behavioral Sciences*, 2, 100033. <https://doi.org/10.1016/j.crbeha.2021.100033>
- Abdelrazek, A., Eid, Y., Gawish, E., Medhat, W., & Hassan, A. (2023). Topic modeling algorithms and applications: A survey. *Information Systems*, 112, 102131. <https://doi.org/10.1016/j.is.2022.102131>
- Alammar, J., & Grootendorst, M. (2024). *Hands-on large language models - Language understanding and generation*. Sebastopol, California, USA: O'Reilly books.
- Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. *International Journal of Advanced Computer Science and Applications*, 6(1). <https://doi.org/10.14569/IJACSA.2015.060121>
- Angelov, D. (2020). Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*. 7. <https://doi.org/10.48550/arXiv.2008.09470>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8, 439-453. https://doi.org/10.1162/tacl_a_00325
- Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and BERTopic to demystify twitter posts. *Frontiers in sociology*, 7, 886498. <https://doi.org/10.3389/fsoc.2022.886498>
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *ArXiv*. <https://doi.org/10.48550/arXiv.2203.05794>
- Hankar, M., Kasri, M., & Beni-Hssane, A. (2025). A comprehensive overview of topic modeling: Techniques, applications and challenges. *Neurocomputing*, 628, 129638. <https://doi.org/10.1016/j.neucom.2025.129638>
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., et al. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169-15211. <https://doi.org/10.1007/s11042-018-6894-4>
- Kern, M., Park, G., Eichstaedt, J., Schwartz, H., Sap, M., Smith, L., et al. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological Methods*, 21(4), 507-525. <https://doi.org/10.1037/met0000091>
- Kherwa, P., & Bansal, P. (2020). Topic modeling: A comprehensive review. *EAI Endorsed Transactions*, 7(24), e2. doi:10.4108/eai.13-7-2018.159623
- Kocak, E., Okumus, F., & Altin, M. (2023). Global pandemic uncertainty, pandemic discussion and visitor behaviour: A comparative tourism demand estimation for the US. *Tourism Economics*, 29(5), 1225-1250. <https://doi.org/10.1177/13548166221100692>
- Leta, S. D., & Chan, I. C. C. (2021). Learn from the past and prepare for the future: A critical assessment of crisis management research in hospitality. *International Journal of Hospitality Management*, 95, 102915. <https://doi.org/10.1016/j.ijhm.2021.102915>

- Liu, L.-W., Pahrudin, P., Tsai, C.-Y., & Hao, L. (2024). Disaster, risk and crises in tourism and hospitality field: A pathway toward tourism and hospitality management framework for resilience and recovery process. *Natural Hazards Research*, 4(4), 653–668. <https://doi.org/10.1016/j.nhres.2024.06.001>
- Ma, L., Chen, R., Ge, W., Rogers, P., Lyn-Cook, B., Hong, et al. (2025). AI-powered topic modeling: comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women. *Experimental Biology and Medicine*, 250, 10389. <https://doi.org/10.3389/ebm.2025.10389>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. New York, NY, USA. Cambridge University Press.
- McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11), 205. doi:10.21105/joss.00205
- McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29), 861. <https://doi.org/10.21105/joss.00861>
- Rejeb, A., Rejeb, K., Simske, S., & Süle, E. (2025). Industry 5.0 research: an approach using co word analysis. *Discover Sustainability*, 6(1), 402. <https://doi.org/10.1007/s43621-025-01252-3>
- Sampaio, C., Farinha, L., Sebastião, J. R., & Fernandes, A. (2023). Tourism industry at times of crisis: a bibliometric approach and research agenda. *Journal of Hospitality and Tourism Insights*, 6(4), 1464–1484. <https://doi.org/10.1108/JHTI-08-2021-0223>
- Sy, C. Y., Maceda, L. L., Flores, N. M., & Abisado, M. B. (2024). Unsupervised machine learning approaches in NLP: A comparative study of topic modeling with BERTopic and LDA. *International Journal of Intelligent Systems and Applications in Engineering* 12(21s), 3276–3283.
- Wijesekara, C., Tittagalla, C., Jayathilaka, A., Ilukpotha, U., Jayathilaka, R., & Jayasinghe, P. (2022). Tourism and economic growth: A global study on Granger causality and wavelet coherence. *PLOS ONE*, 17(9), e0274386. <https://doi.org/10.1371/journal.pone.0274386>
- Wong, A. K. F., Kim, S., Liu, Y. Y., & Grace Baah, N. (2025). COVID-19 research in hospitality and tourism: critical analysis, reflection, and lessons learned. *Journal of Hospitality & Tourism Research*, 49(1), 188–204. <https://doi.org/10.1177/10963480231156079>
- Zhong, L., Sun, S., Law, R., & Li, X. (2021). Tourism crisis management: evidence from COVID-19. *Current Issues in Tourism*, 24(19), 2671–2682. <https://doi.org/10.1080/13683500.2021.1901866>