Topic modeling in tourism research: a bibliometric study

Valentinus Roby Hananto[1,2], Vivine Nurcahyawati[2], Tutut Wurijanto[2], Titik Lusiani[2], Mate Kovacs[1]
[1]College of Information Science and Engineering, Ritsumeikan University, Ibaraki, Japan
[2]Department of Information Systems, Faculty of Technology and Informatics, Universitas Dinamika, Surabaya, Indonesia

Article Info

Article history:

Received Mar 4, 2025
Revised Mar 6, 2026
Accepted May 26, 2026

Keywords:

Bibliometric
Citation analysis
Research trend
Smart tourism
Topic modeling

ABSTRACT

The application of topic modeling in the tourism domain has become a popular research topic in the last decade. This study aims to provide a comprehensive bibliometric analysis of topic modeling research in tourism over the period 2010 to 2023. The data for this study were sourced from the Scopus database, a widely recognized repository of peer-reviewed literature. The search was restricted to publications published from January 1, 2010, to December 31, 2023, to capture the evolution and current state of this rapidly growing field. Using VOSviewer and SciMAT software to analyze articles in the Scopus database, the study identified key trends, influential authors, and future research directions. This study indicates the growth and development in topic modeling for tourism research, with more than 100 Scopus-indexed papers published annually in 2023 alone. The results of this study show that topic modeling has a wide range of applications in tourism, demonstrating its utility in various contexts to understand tourist behavior and enhance smart tourism initiatives.

Corresponding Author:

Valentinus Roby Hananto
College of Information Science and Engineering, Ritsumeikan University
Ibaraki, Osaka, Japan
Email: hananto@fc.ritsumei.ac.jp

1. INTRODUCTION

The advent of smart tourism has transformed the way tourist destinations are managed and experienced. Smart tourism refers to the integration of advanced technologies—such as the internet of things (IoT), big data analytics, artificial intelligence, and mobile technologies—into tourism ecosystems to improve efficiency, sustainability, and personalization of services [1]. It leverages real-time data from interconnected sources, such as social media platforms, sensors, travel apps, and government databases, to create seamless and enriched experiences for tourists while optimizing resource usage for service providers [2]. Smart tourism has become increasingly important as it meets the evolving demands of modern travelers, who seek not only convenience and accessibility but also highly customized, meaningful experiences. Implementing smart technologies enables stakeholders to cater more efficiently to tourist requirements, enhance operational efficiency, control crowd flow, and minimize environmental effects, thus promoting the development of sustainable tourism [3]. In the post-pandemic era, smart tourism also plays a crucial role in ensuring health and safety by facilitating contactless services, monitoring visitor flows, and adapting opportunities in real-time to changing conditions.

With the propagation of these digital technologies, tourism research has increasingly embraced data-driven methods to extract actionable insights from the vast amounts of unstructured data generated across various platforms. A significant portion of this data comes from user-generated content (UGC)—such as online reviews, social media posts, and blogs—where tourists freely share experiences, express opinions, and provide feedback. Among emerging technologies, topic modeling has gained substantial traction. Topic modeling is a machine learning method that identifies latent themes or topics within large text corpora without the need for prior annotations [4]. It helps uncover patterns and trends that might otherwise remain hidden, offering valuable insights for tourism stakeholders. By analyzing UGC through topic modeling, researchers and practitioners can better understand tourists' preferences, predict trends, and make informed decisions to enhance service quality and destination management.

The development of topic modeling methods has evolved over the years, contributing uniquely to the analysis of unstructured data in tourism research. These methods have enabled researchers to uncover latent themes, providing deeper insights into tourist behavior and preferences. One of the early methods used in this domain is Latent Semantic Analysis (LSA), which emerged in the late 1990s [5]. LSA applies singular value decomposition to identify and represent the underlying structure of relationships among terms and documents. This method has been utilized in tourism research to analyze large datasets of text, such as online reviews and social media posts, uncovering insights into consumer sentiments and preferences [6]. The latent Dirichlet allocation (LDA), introduced by Blei *et al.* [4], has been one of the most popular topic modeling methods and has been widely adopted in tourism research since the early 2010s. LDA assumes that each document is a mixture of topics, and each topic is characterized by a distribution of words. In the tourism domain, LDA has been extensively used to analyze reviews from various sources, including tourist destinations [7], hotel reviews [8], and restaurant reviews [9]. Building on LDA, the dynamic topic model (DTM) [10] gained traction in tourism research around 2015. DTM captures the evolution of topics over time, making it ideal for analyzing temporal shifts in consumer preferences. For example, Xu *et al.* [11] applied DTM to community-contributed geotagged photos (CCGPs) to obtain temporally fine-grained topic distributions of users and locations. Around the same period, non-negative matrix factorization (NMF) emerged as an alternative to LDA. Unlike LDA, which is probabilistic, NMF decomposes the word-document matrix into non-negative components, producing interpretable topics. A study by Jansen *et al.* [12] employed NMF to perform market segmentation by analyzing online customer patterns over time. In the early 2010s, the correlated topic model (CTM) became increasingly popular. CTM, an extension of LDA, accounts for the correlations between topics, making it suitable for the analysis of interdependent topics [13]. For example, a recent study employed CTM to capture correlations between latent topics from TripAdvisor reviews [14]. Their proposed model aimed to predict optimal travel destinations based on user preferences by combining sentiment scores and star ratings. The hierarchical Dirichlet process (HDP), developed by Teh *et al.* [15], has been applied to heterogeneous tourism datasets. HDP can automatically determine the optimal number of topics, addressing one of the limitations of LDA. A recent study by Zhang *et al.* [16] categorizes documents into 22 sub-domain categories, including tourism. Their work effectively discovered inter-topic correlations and differences among subtopics in the topic hierarchy.

By the early 2020s, advances in natural language processing (NLP) led to the adoption of BERT-based Topic Models. BERT (Bidirectional Encoder Representations from Transformers), introduced by Devlin *et al.* [17], captures contextual word embeddings, making it effective for analyzing short texts. Rey-Moreno *et al.* [18] used a BERT-based topic model (BERTopic) to analyze hotel guest reviews, revealing latent themes relevant to guest satisfaction and trust. These findings aided in the formulation of marketing strategies and enhanced comprehension of the hospitality sector. The progression from early probabilistic models like LDA to more sophisticated models such as HDP and BERTopic highlights the increasing complexity and volume of data in smart tourism. These models have empowered researchers and practitioners to extract data-driven insights, to enhance service offerings, and to create personalized tourism experiences.

The exploration of topic modeling within tourism research has gained traction in recent years; however, focused bibliometric studies analyzing this specific area remain scarce. While some previous reviews [1], [19], [20] have examined broader themes in smart tourism, they have not delved into the unique contributions and trends of topic modeling. Although topic modeling advancements have progressed in smart tourism research, the evolution of its application over time and the impact of various research clusters on its development have not been extensively explored. A dedicated bibliometric study could provide a comprehensive overview of key themes, leading authors, and influential publication sources within the realm of topic modeling in tourism research. A thorough analysis would increase the prominence of this emerging field and also provide direction for incoming researchers aiming to identify potential areas of study. Additionally, it could facilitate interdisciplinary collaboration, fostering a more integrated approach to studying tourism dynamics.

This study aims to fill this gap by conducting a comprehensive bibliometric analysis of topic modeling in tourism research from 2010 to 2023. By employing metadata from the Scopus database, this research utilizes VOSviewer to visualize bibliometric networks and explores thematic evolution with SciMAT, in order to uncover significant trends, prominent authors, and potential future research directions. The specific objectives of this study are to (i) analyze the types of documents, sources, authors, affiliations, and countries contributing to topic modeling in tourism over the last decade, (ii) identify the top journals and sources publishing on topic modeling usage within the tourism domain, (iii) determine the most influential articles, authors, and institutions shaping this field, (iv) uncover key themes and trends to understand the thematic landscape within this field; (v) provide insights into potential future directions for research and practical applications of topic modeling in tourism. Through these objectives, this study aims to offer a detailed overview of the current state of topic modeling in tourism research and serve as a valuable resource for researchers and practitioners looking to explore or expand upon this evolving field.

2. METHOD

This section presents the research design and methodology employed to perform a comprehensive bibliometric analysis of topic modeling within tourism research from 2010 to 2023. The study follows a systematic approach, organized into five key phases: data collection, data preprocessing, visualization, and bibliometric analysis. Each phase ensures a rigorous exploration of trends, key themes, and research dynamics within the field.

2.1. Data collection

The data for this study was sourced from the Scopus database, one of the most extensive and widely recognized repositories of peer-reviewed literature. Scopus was chosen for its comprehensive coverage of high-impact journals, conference proceedings, and other scholarly sources relevant to the intersection of technology and tourism. A structured search strategy was employed to identify relevant publications. The search terms combined keywords related to topic modeling technologies (e.g., "topic modeling," "Latent Dirichlet Allocation," "BERTopic") and terms associated with the smart tourism domain (e.g., "tourism," "smart tourism," "accommodation," "hospitality," "hotel," "culinary," "restaurant," "attraction"). The search was restricted to publications from January 1, 2010, to December 31, 2023, to capture the evolution of topic modeling within the tourism context over the last decade. The data collection was conducted on August 15, 2024, and the initial search yielded 637 publications, with the search query as follows: (TITLE-ABS-KEY (”topic modeling” OR ”latent Dirichlet allocation” OR ”latent semantic analysis” OR ”BERTopic”) AND TITLE-ABS-KEY (”tourism” OR ”tourist” OR ”accommodation” OR ”hospitality” OR ”hotel” OR ”culinary” OR ”restaurant” OR ”attraction”) AND PUBYEAR > 2009 AND PUBYEAR < 2024).

2.2. Data preprocessing

To ensure relevance, inclusion and exclusion criteria were applied. Only peer-reviewed journal articles, conference papers, and review articles written in English were considered, as these sources typically represent rigorous and high-quality research in the field. Publications that were not directly related to tourism or that did not explicitly use topic modeling as a machine-learning method were excluded through manual screening. After applying these criteria, 522 documents were retained for final analysis. The filtered data was then preprocessed to ensure accuracy and consistency. This involved the normalization of keywords to account for variations in spelling and formatting. For example, ”smart tourism” and ”intelligent tourism” were standardized under a single term. Next, a keyword co-occurrence analysis was performed to identify the most frequently occurring terms within the dataset. This step was crucial in understanding the thematic landscape of topic modeling in tourism research and in preparing the data for further analysis.

2.3. Bibliometric analysis

The bibliometric analysis was conducted using two specialized software tools: VOSviewer and SciMAT. VOSviewer's strength lies in its ability to handle large datasets and produce detailed visual maps that reveal the structure of research fields. VOSviewer was used to create and visualize keyword co-occurrence networks. Keyword co-occurrence networks highlight the most prominent themes in the dataset by showing how frequently keywords appeared together in the publications. Clusters of related keywords were identified, providing insights into the main topics of interest within the tourism research community. Citation analysis was also conducted to determine the most cited articles and authors. This analysis helped to identify seminal works and influential contributors to the field of topic modeling in tourism. SciMAT was employed to perform a longitudinal analysis of the research themes. This tool enabled the study to track the evolution of topics over time, revealing how the focus of topic modeling in tourism research has shifted from 2010 to 2023. Thematic evolution maps were created to illustrate the progression of research themes over the study period. These maps revealed how certain topics gained prominence or declined in relevance over time.

3. RESULTS AND DISCUSSION

This section presents the findings from the bibliometric analysis of topic modeling in tourism research over the period from 2010 to 2023. This section is organized into five subsections: publication trends, citation analysis, keyword analysis, thematic evolution, and discussion. Each subsection highlights key insights into the development and focus areas of topic modeling within the tourism research community.

3.1. Publications trends

The analysis of publication trends highlights a consistent rise in the use of topic modeling in tourism research from 2010 to 2023. Figure 1 shows that the number of publications were relatively low from 2010 to 2015 (less than 20 publications annually), with minor fluctuations. However, starting in 2016, there is a noticeable increase, followed by a rapid and sustained growth after 2019. This surge indicates the growing recognition of topic modeling techniques as valuable tools in understanding tourism trends, consumer behavior, and market dynamics. As shown in Figure 1, the annual number of publications rose sharply from approximately 20 in 2016 to over 130 by 2023. The steep upward trend from 2020 onward reflects a broader adoption of digital data analytics and smart tourism technologies. The COVID-19 pandemic likely acted as a catalyst during 2020-2021, prompting researchers to explore new methods such as text mining and topic modeling due to the increased reliance on digital platforms for tourism services and communication. This growth trajectory suggests that the field will continue to expand as the tourism industry integrates more data-driven decision-making processes, leveraging online reviews, social media, and other forms of user-generated content to enhance tourism management and customer satisfaction. Regarding types of publications during the whole period, there are 317 journal articles, 179 conference papers, 14 review papers, and 12 book chapters and letters. The analysis reveals a wide array of sources influencing the field. Table 1 displays the six leading journals by publication volume, collectively responsible for 25% of all publications, highlighting their significant role in shaping discussions on topic modeling in tourism research.

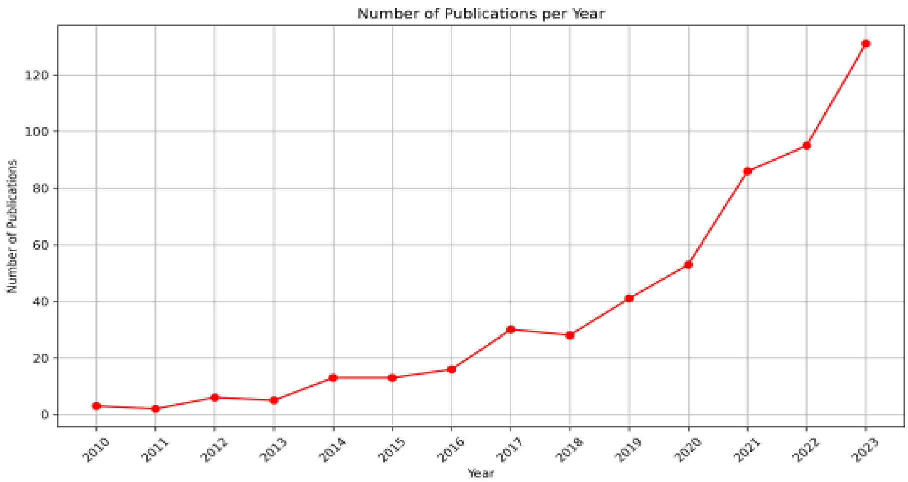


Figure 1. Number of publications from 2010 to 2023

Table 1. The top journals in terms of publication volume during 2010-2023

Rank	Journal title	Frequency	Subject area	Publisher
1	Sustainability	23	Computer science; energy; environmental science; social sciences	MDPI
2	International Journal of Contemporary Hospitality Management	15	Business, Management and Accounting	Elsevier
3	International Journal of Hospitality Management	12	Business, Management and Accounting	Elsevier
4	Current Issues in Tourism	12	Business, Management and Accounting; Social Sciences	Taylor and Francis
5	Tourism Management	8	Business, Management and Accounting; Social Sciences	Elsevier
6	Journal of Hospitality and Tourism Technology	8	Business, Management and Accounting; Computer Science	Emerald

3.2. Citation analysis

This study conducted a citation analysis to identify the most influential works and authors in tourism research. Using Scopus, the top-cited papers from 2010 to 2023 were identified, and Scimago Journal Rank (SJR) score was utilized to assess the impact of the journals that published these works (see Table 2). The SJR data (collected on October 30, 2024) provides insight into the quality and influence of the journals, complementing the citation analysis by highlighting the academic standing of the outlets.

Table 2. The most cited journal papers during 2010-2023

Rank	Title	Citation	Year	Journal (SJR)
1	Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent Dirichlet allocation [21]	708	2017	Tourism management (3.35)
2	Green, circular, bioeconomy: a comparative analysis of sustainability avenues [22]	659	2017	Journal of cleaner production (2.06)
3	The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: a text mining approach [23]	319	2016	International journal of hospitality management (2.92)
4	Analysing tripadvisor reviews of tourist attractions in Phuket, Thailand [24]	225	2019	Tourism management (3.35)
5	Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews [25]	217	2019	Tourism management (3.35)
6	Sentiment classification of consumer-generated online reviews using topic modeling [26]	178	2017	Journal of hospitality marketing and management (3.35)
7	Business intelligence in online customer textual reviews: understanding consumer perceptions and influential factors [8]	174	2017	International journal of information management (5.78)
8	W2VLDA: almost unsupervised system for aspect based sentiment analysis [27]	164	2018	Expert systems with applications (1.88)
9	Social media analytics and value creation in urban smart tourism ecosystems [28]	150	2017	Information and management (2.59)
10	Motivation and satisfaction of Chinese and U.S. tourists in restaurants: a cross-cultural text mining of online reviews [29]	149	2020	Tourism management (3.35)

The results show that the article by Guo *et al.* [21] is the most cited paper with 708 citations. Published in Tourism Management (SJR: 3.35), this study is pivotal in applying LDA to tourism research, establishing a widely adopted methodology for analyzing online reviews and tourist satisfaction. The work by D'Amato *et al.* [22], ranking as the second most referenced (659 citations), appeared in the Journal of Cleaner Production (SJR: 2.06). It delves into sustainability by examining green, circular, and bioeconomies. This publication underscores the increasing emphasis on sustainable tourism and highlights the shift towards making environmental sustainability a central issue in the industry.

Several other high-impact papers are published in journals with strong SJR scores, emphasizing the quality of research in tourism and related domains. Tourism Management Journal (SJR: 3.35) appears multiple times in the list, hosting influential works such as the study on analyzing TripAdvisor reviews [24] with 225 citations, the study of importance-performance analysis (IPA) on tourism reviews by Bi *et al.* [25] with 217 citations, and the study on restaurant customers' satisfaction [29]. This underscores the significant role of the journal in advancing tourism analytics and data-driven research. The highest SJR-scored journal in the list is the International Journal of Information Management (SJR: 5.78), which published a study about business intelligence in online customer reviews by Xu *et al.* [8], gathering 174 citations. This reflects the increasing interdisciplinary nature of tourism research, where insights from information management contribute to studying consumer behavior.

An interesting trend is the focus on topic modeling and sentiment analysis. Several related papers from the top journals are the study by Calheiros *et al.* [26] with 178 citations, published in the Journal of Hospitality Marketing and Management (SJR: 3.35), and a study by Garcia-Pablos *et al.* [27] with 164 citations, appearing in Expert Systems with Applications (SJR: 1.88). These contributions highlight the growing importance of machine learning techniques and NLP in tourism analytics, helping researchers and practitioners analyze large volumes of unstructured data from online platforms.

The distribution of SJR scores among these journals suggests a balanced mix of tourism-specific and interdisciplinary journals, indicating that smart tourism research contributes to multiple fields, including management, sustainability, and information science. The repeated appearance of journals like Tourism Management and International Journal of Hospitality Management demonstrates the central role of these journals in shaping the research landscape, while the presence of high-SJR interdisciplinary journals reflects the broadening scope of tourism research.

The next analysis focuses on the publications based on the research subject area (based on the Scopus database metadata). Table 3 provides an analysis of publication trends from 2010 to 2023, with a focus on the most cited papers published in 2022-2023. Recently, "computer science" has been the leading field with 116 papers published in the last 2 years, followed by "business, management, and accounting" (96 papers), and "social sciences" (82 papers). 116 papers were published in the "Computer Science" subject area in 2022 and 2023. Some of them implemented topic modeling on online reviews on tourist attractions [30], accommodation [31]-[34], and restaurants [35], [36]. About 96 papers were published in the "business, management, and accounting" subject area in 2022-2023. This work sheds new light on the emergence of a body of research at the intersection of hospitality and tourism management and data science [20], [37], [38]. Meanwhile, recent papers that were published in the "Social Sciences" mainly discussed the COVID-19 pandemic [9], [39], [40] and tourist destination branding [41]-[43]. "engineering" and "decision sciences" subject area, while having fewer publications, still produce impactful work as indicated by their recent papers [44], [45]. This citation analysis suggests a strong emphasis on interdisciplinary approaches, with a notable overlap between computer science and other domains such as business and social sciences.

Table 3. Research subject area

Research subject area	Total papers (2010-2023)	Total papers (2022-2023)	Most cited papers (2022-2023)
Computer science	363	116	Venugopalan and Gupta [35]; Lee [31]; Puri *et al.* [46]; Alsayat [47]
Business, management, and accounting	189	96	Mariani and Baggio [20]; Venugopalan and Gupta [35]; Lee [31]; Nilashi *et al.* [48]
Social sciences	162	82	Zibarzani *et al.* [9]; Nunkoo *et al.* [49]; Mehta *et al.* [50]; Puri *et al.* [46]
Engineering	132	49	Mishra *et al.* [44]; Afaq *et al.* [45]; Vaish *et al.* [51]; Luo *et al.* [52]
Decision sciences	86	37	Venugopalan and Gupta [35]; Lee [31] Mishra *et al.* [44]; Afaq *et al.* [45];

3.3. Keyword analysis

The keyword co-occurrence analysis revealed several recurring themes and topics within the literature. Figure 2 presents keyword co-occurrence network visualization provided by the VOSviewer. The research on topic modeling in the tourism domain can be analyzed based on the identified clusters and connections between keywords. The size of the nodes reflects the frequency of each keyword, while the thickness of the edges indicates the strength of their association.

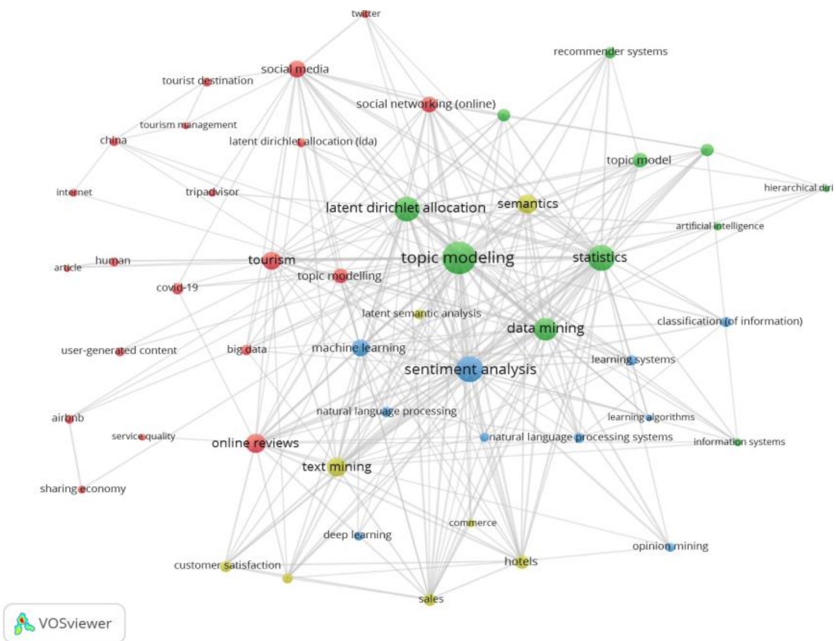


Figure 2. VOSviewer keyword co-occurrence network

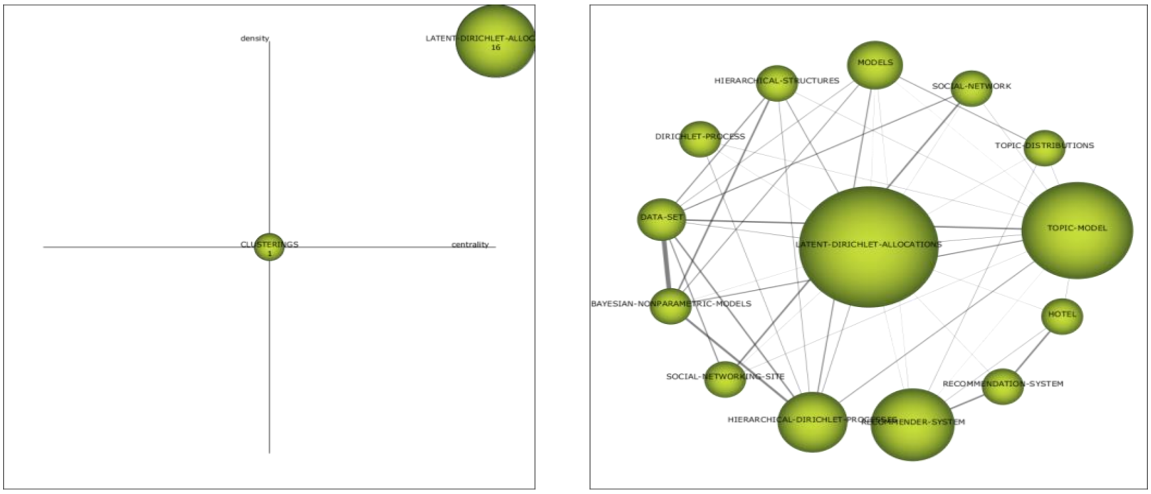
There are four clusters identified in the keyword co-occurrence network. Cluster 1 (red) emphasizes the role of social networks in smart tourism and destination image analysis. Top keywords in this cluster are social media, tourism, online reviews, user-generated content, COVID-19, Airbnb, TripAdvisor, tourist destination, customer satisfaction, and service quality. The next three clusters are closely related to each other, mainly covering the methods and technology adoption in the smart tourism domain. The main theme of Cluster 2 (green) is about recommender systems and Artificial Intelligence (AI) for smart tourism, with top keywords such as topic modeling, latent Dirichlet allocation, data mining, recommender systems, and hierarchical Dirichlet process. This cluster focuses on advanced AI and machine learning methods, such as LDA, and their applications in recommender systems within the tourism context. Cluster 3 (blue) is mainly about sentiment analysis and text mining. With the top keywords of sentiment analysis, NLP, opinion mining, and deep learning, this cluster highlight the application of text mining and NLP technologies for evaluating tourist opinions. Lastly, Cluster 4 (yellow) representing the theme of statistical and data-driven approaches to analyze semantics of large-scale textual data. Top keywords from this cluster are semantics, latent semantic analysis, text mining, and customer satisfaction. Central terms such as ”topic modeling,” ”latent Dirichlet allocation,” ”sentiment analysis,” and ”text mining” serve as bridges between the clusters, indicating their importance in the tourism topic modeling field. Meanwhile, the presence of terms like ”COVID-19,” ”online reviews,” and ”artificial intelligence” suggests growing interest in adapting methodologies to current events and technological advancements. The emphasis on recommender systems, sentiment analysis, and customer satisfaction reflects practical applications in improving tourism services and understanding consumer behavior.

3.4. Thematic evolution

The thematic evolution of topic modeling in tourism research was examined using SciMAT, revealing how research themes have developed and shifted over the 2010-2023 period.

3.4.1. Early period (2010-2014)

The early period of topic modeling in tourism was characterized by exploratory studies focused on applying LDA to datasets primarily from UGC, as researchers began to recognize the potential of topic modeling for analyzing tourism reviews. Figure 3 shows the strategic diagram, a two-dimensional graph that plots clusters or themes based on their centrality and density, and clusters' network that visualizes relationships between themes. Figure 3(a) presents the strategic diagram for the early period (2010–2014), plotting themes based on their centrality and density. Figure 3(b) shows the cluster's network, visualizing the relationships and proportions between identified themes. The theme latent-Dirichlet-allocations is in the upper right quadrant of the strategic diagram (high centrality and high density), which is interpreted as a core theme driving the research field. In the cluster's network, notable clusters that have a high proportion of the documents in this period are latent-Dirichlet-allocations (16 publications), topic-model (12), recommender-system (8), and hierarchical-Dirichlet-processes (6).

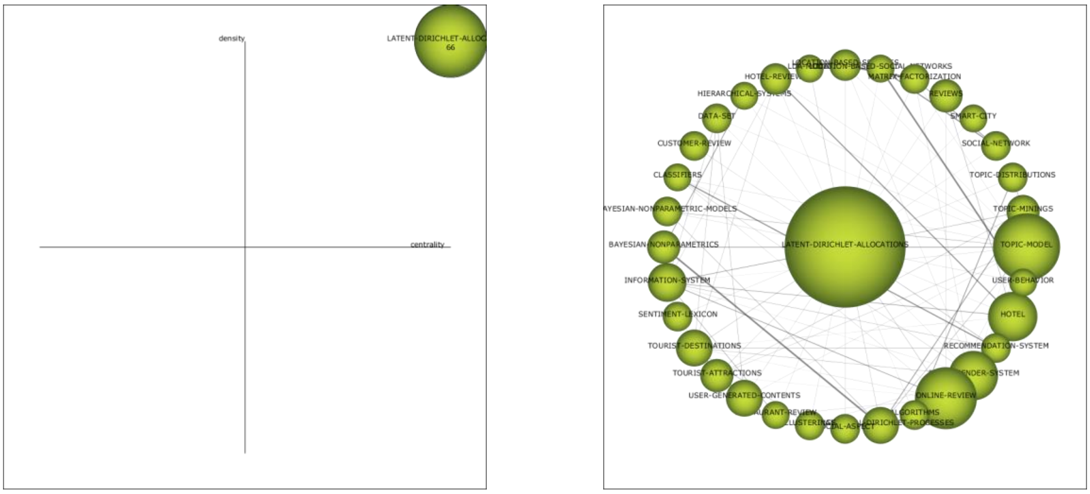


(a) (b)

Figure 3. Strategic diagram and cluster's network of early period (2010–2014); (a) theme centrality and density and (b) inter-theme relationships

3.4.2. Progressive period (2015-2019)

This period saw the expansion of topic modeling applications, with an increasing focus on smart tourism and big data. Thematic clusters related to smart tourism such as ”online review” and ”hotel” emerged as key areas of interest, reflecting the growing integration of technology in the domain. Research during this period also began to explore the scalability of topic modeling to larger datasets and more complex research questions. In the strategic diagram and cluster's network (Figure 4), “latent-Dirichlet-allocationS” is still the well-developed theme with strong connection to other clusters. However, the proportion of the clusters are a little bit shifted, with the emerging of “online-review” as one of the notable themes. Figure 4(a) presents the strategic diagram for the progressive period (2015–2019), illustrating this shift in theme prominence. Figure 4(b) shows the cluster's network for this period, highlighting the growing role of user-generated content. This mark the rising popularity of user-generated content that was utilized more in this period. In this period, the most productive themes are latent-Dirichlet-allocations (54 publications), topic-model (24), online-review (21), hotel (14), and recommender-system (14).

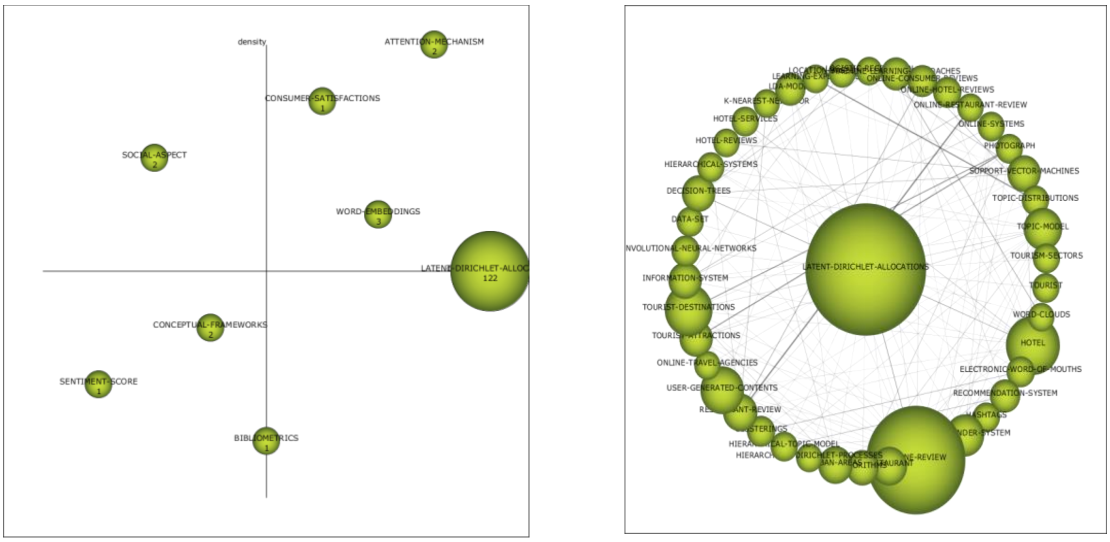


(a) (b)

Figure 4. Strategic diagram and cluster's network of progressive period (2015–2019); (a) theme centrality and density and (b) inter-theme relationships

3.4.3. Mature period (2020-2023)

This recent period has been influenced by the COVID-19 pandemic, which accelerated the adoption of digital tools in tourism practices. This period also witnessed a surge in interdisciplinary research, combining topic modeling with other analytical techniques to address emerging challenges. ”LATENT-DIRICHLET-ALLOCATIONS” still has the highest centrality, but not the highest density anymore (see Figure 5). Figure 5(a) presents the strategic diagram for the mature period (2020–2023), indicating the redistribution of centrality and density with several clusters appearing in the top-right quadrant. Figure 5(b) shows the cluster's network for this period, reflecting the interdisciplinary growth and the influence of advanced NLP techniques such as attention mechanisms and word embeddings. Several clusters appear in the top-right quadrant (“attention-mechanism”, “consumer-satisfactions”, and “word-embeddings”), suggesting a promising area to explore the technology of the large language model. Meanwhile, some clusters are in the lower-left quadrant (e.g. “sentiment score”) suggesting the maturity of these themes. The most productive themes in this period are latent-Dirichlet-allocations (94), online-review (73), hotel (29), tourist-destinations (23), and user-generated-contents (19).



(a) (b)

Figure 5. Strategic diagram and cluster's network of mature period (2020–2023); (a) theme centrality and density and (b) inter-theme relationships

3.5. Discussion

The application of topic modeling in tourism research has gained significant attention over the past decade, particularly in understanding customer satisfaction, sentiment analysis, and business intelligence. The following review synthesizes key findings from top-10 cited studies published between 2010 and 2023 (see Table 2), which have employed topic modeling techniques to analyze tourism-related textual data.

3.5.1. Tourist satisfaction and sentiment analysis

A study by Guo *et al.* [21] applies LDA to analyze tourist satisfaction from online ratings and reviews. The study demonstrates the effectiveness of topic modeling in identifying key themes influencing satisfaction and dissatisfaction, providing valuable insights for tourism management. Similarly, Xu and Li [23] investigate customer satisfaction and dissatisfaction antecedents across various hotel types using a text mining approach, revealing critical service attributes that impact consumer experiences. Another prominent study by Calheiros *et al.* [26] focuses on the sentiment classification of online reviews, integrating topic modeling techniques to extract insights from unstructured text data. This approach enhances the ability to categorize sentiments, aiding businesses in tailoring their services accordingly. Furthermore, a study by Jia [29] explores cross-cultural differences in tourist motivation and satisfaction, using text mining to compare the restaurant experiences of Chinese and US tourists.

3.5.2. Tourism business intelligence and consumer perception

The role of business intelligence in tourism research is highlighted by Mehta *et al.* [50], where topic modeling is utilized to analyze online customer textual reviews. This study identifies consumer perceptions and influential factors affecting purchasing decisions. The findings emphasize the importance of leveraging big data analytics in strategic business decision-making. A related study by Bi *et al.* [25] employs importance-performance analysis (IPA) on online reviews to assess tourist attractions, illustrating the value of the wisdom of crowds in enhancing service offerings. Similarly, Taecharungroj and Mathayomchan [24] applies topic modeling to analyze TripAdvisor reviews of tourist attractions in Phuket, Thailand, uncovering patterns that influence visitor satisfaction and preferences.

3.5.3. Sustainable tourism

With growing concerns over sustainability in tourism, D'Amato *et al.* [22] present a comparative analysis of green, circular, and bio economies, demonstrating how topic modeling can uncover sustainability avenues in tourism. This study contributes to the broader discourse on sustainable tourism management. Furthermore, Brandt *et al.* [28] explore the role of social media analytics in urban smart tourism ecosystems, utilizing topic modeling to understand value creation in digital tourism environments. This research underscores the transformative impact of digital technologies in shaping smart tourism practices.

3.5.4. Advanced methodologies in topic modeling

Several studies have explored methodological advancements in topic modeling for tourism research. For instance, Garcia-Pablos *et al.* [27] introduce W2VLDA, an unsupervised system for aspect-based sentiment analysis, enhancing the accuracy and interpretability of sentiment classification in tourism reviews. Likewise, other studies have demonstrated methods to refine topic modeling, contributing to the growing sophistication of computational approaches in tourism analytics [35].

Citation and keyword analyses identify several influential works and emerging trends within the field. Highly cited studies have laid the foundation for the application of topic modeling in tourism, particularly in UGC analysis. These works have demonstrated the benefits of topic modeling, contributing to its adoption in academic and industry settings. As tourism becomes increasingly intertwined with technology, the ability to analyze digital data through topic modeling has become essential. This trend is particularly evident in the rise of studies focusing on smart tourism, where topic modeling is used to analyze big data, optimize tourism experiences, and enhance the sustainability of tourism operations. Another emerging trend is the use of topic modeling to assess the impacts of global crises on tourism. The COVID-19 pandemic, in particular, has spurred a wave of research aimed at understanding how the industry has been affected and how it can recover. Topic modeling has been employed to analyze changes in tourist behavior, sentiment, and preferences during and after the pandemic, providing valuable insights for crisis management and recovery strategies.

Despite its growing popularity, the application of topic modeling in tourism research is not without challenges. One of the primary limitations is the quality of the data used. Topic modeling relies on large datasets, often derived from UGC, which can be noisy and unstructured. This introduces the risk of bias and misinterpretation of the results, especially if the data is not properly cleaned or if the model parameters are not carefully selected. Additionally, topic modeling, particularly methods like LDA, can sometimes produce ambiguous or incoherent topics that require expert interpretation. This subjective element can affect the reproducibility and reliability of the results, making it difficult to draw definitive conclusions. Another challenge is the integration of topic modeling with other analytical techniques. While interdisciplinary approaches are on the rise, there is still a need for more robust frameworks that combine topic modeling with methods such as sentiment analysis, network analysis, and machine learning. Such integrations can enhance the depth and breadth of insights derived from the data but require advanced technical expertise and computational resources.

Based on the findings and the identified challenges, several future research directions are as follows: i) enhanced data quality and preprocessing: future research should focus on improving data quality through better preprocessing techniques, such as advanced text cleaning, normalization, and noise reduction. This will help minimize biases and improve the accuracy of topic modeling; ii) development of new topic modeling techniques: there is potential for developing new or improved topic modeling algorithms that address the limitations of current methods, such as the production of ambiguous topics. Methods that incorporate contextual information or semantic understanding could lead to more coherent and interpretable topics; iii) integration with other analytical methods: as interdisciplinary research becomes more prevalent, there is a need to develop frameworks that seamlessly integrate topic modeling with other analytical methods. For example, combining topic modeling with sentiment analysis could provide a more comprehensive understanding of tourist emotions and experiences; iv) application to emerging issues in tourism: the adaptability of topic modeling makes it a suitable tool for addressing emerging issues in tourism, such as the impact of climate change, the rise of virtual tourism, and the evolution of tourist behavior in a post-pandemic world. Research should explore how topic modeling can be applied to provide insight that can support policy to address emerging issues.

4. CONCLUSION

This study has provided a comprehensive bibliometric analysis of topic modeling in tourism research over the period from 2010 to 2023. By examining publication trends, citation networks, keyword co-occurrence, and thematic evolution, we have gained valuable insights into the development and current state of this rapidly growing field. The integration of topic modeling into tourism research has significantly improved the understanding of customer behavior, business intelligence, and tourism sustainability. The collected studies illustrate how advanced topic modeling techniques can extract meaningful insights from large-scale textual data, providing valuable implications for academic and industry practitioners. There has been notable growth in the field since 2019, highlighting the increasing importance of data-driven methods in tourism research. Although earlier efforts concentrated on destination image analysis, the use of topic modeling has expanded to include a diverse range of subjects, such as tourist behavior, smart tourism, and crisis management. Some highly cited papers have profoundly influenced the field, establishing foundational insights for future research. These pioneering studies have showcased the practical benefits of topic modeling, particularly in examining UGC, and have inspired continued innovation and inquiry.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Valentinus Roby Hananto	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Vivine Nurcahyawati		✓		✓		✓		✓	✓		✓			✓
Tutut Wurijanto		✓		✓		✓		✓	✓		✓			✓
Titik Lusiani		✓		✓		✓		✓	✓		✓			✓
Mate Kovacs	✓			✓		✓			✓	✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known financial, personal, or professional conflict of interests that could have influenced the work reported in this paper.

DATA AVAILABILITY

The data supporting the findings of this study were obtained from primary data collected from the Scopus database. The data are available from the corresponding author upon reasonable request.

REFERENCES

[1] U. Gretzel, M. Sigala, Z. Xiang, and C. Koo, "Smart tourism: foundations and developments," *Electronic Markets*, vol. 25, no. 3, pp. 179–188, Sep. 2015, doi: 10.1007/s12525-015-0196-8.

[2] D. Buhalis and A. Amaranggana, "Smart tourism destinations enhancing tourism experience through personalisation of services," in *Information and Communication Technologies in Tourism 2015*, vol. 15, no. 3, I. Tussyadiah and A. Inversini, Eds., Cham: Springer, 2015, pp. 377–389. doi: 10.1007/978-3-319-14343-9_28.

[3] S. Shafiee, A. Rajabzadeh Ghatari, A. Hasanzadeh, and S. Jahanyan, "Developing a model for sustainable smart tourism destinations: a systematic review," *Tourism Management Perspectives*, vol. 31, pp. 287–300, 2019, doi: 10.1016/j.tmp.2019.06.002.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, no. 4–5, pp. 993–1022, 2003.

[5] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse Processes*, vol. 25, no. 2–3, pp. 259–284, Jan. 1998, doi: 10.1080/01638539809545028.

[6] W. Luo, F. Zhuang, Q. He, and Z. Shi, "Quad-tuple PLSA: incorporating entity and its rating in aspect identification," in *Advances in Knowledge Discovery and Data Mining. PAKDD 2012. Lecture Notes in Computer Science*, vol. 7301 LNAI, no. PART 1, P. Tan, S. Chawla, C. K. Ho, and J. Bailey, Eds., Berlin, Heidelberg: Springer, 2012, pp. 392–404. doi: 10.1007/978-3-642-30217-6_33.

[7] T. Ali, B. Marc, B. Omar, K. Soulaimane, and S. Larbi, "Exploring destination's negative e-reputation using aspect-based sentiment analysis approach: case of Marrakech destination on TripAdvisor," *Tourism Management Perspectives*, vol. 40, p. 100892, Oct. 2021, doi: 10.1016/j.tmp.2021.100892.

[8] X. Xu, X. Wang, Y. Li, and M. Haghighi, "Business intelligence in online customer textual reviews: understanding consumer perceptions and influential factors," *International Journal of Information Management*, vol. 37, no. 6, pp. 673–683, Dec. 2017, doi: 10.1016/j.ijinfomgt.2017.06.004.

[9] M. Zibarzani *et al.*, "Customer satisfaction with restaurants service quality during COVID-19 outbreak: a two-stage methodology," *Technology in Society*, vol. 70, p. 101977, Aug. 2022, doi: 10.1016/j.techsoc.2022.101977.

[10] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proceedings of the 23rd international conference on Machine learning - ICML '06*, New York, New York, USA: ACM Press, 2006, pp. 113–120. doi: 10.1145/1143844.1143859.

[11] Z. Xu, L. Chen, Y. Dai, and G. Chen, "A dynamic topic model and matrix factorization-based travel recommendation method exploiting ubiquitous data," *IEEE Transactions on Multimedia*, vol. 19, no. 8, pp. 1933–1945, Aug. 2017, doi: 10.1109/TMM.2017.2688928.

[12] B. J. Jansen, S. Jung, S. A. Chowdhury, and J. Salminen, "Persona analytics: analyzing the stability of online segments and content interests over time using non-negative matrix factorization," *Expert Systems with Applications*, vol. 185, p. 115611, Dec. 2021, doi: 10.1016/j.eswa.2021.115611.

[13] D. M. Blei and J. D. Lafferty, "Correlated topic models," in *Proceedings of the 19th International Conference on Neural Information Processing Systems*, 2006, pp. 147–155.

[14] N. Kumar and B. R. Hanji, "Combined sentiment score and star rating analysis of travel destination prediction based on user preference using morphological linear neural network model with correlated topic modelling approach," *Multimedia Tools and Applications*, vol. 83, no. 22, pp. 61347–61378, Jan. 2024, doi: 10.1007/s11042-023-17995-y.

[15] Y. W. Teh, D. Newman, and M. Welling, "A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation," in *Proceedings of the 20th International Conference on Neural Information Processing Systems*, 2007, pp. 1353–1360.

[16] Y. Zhang, C. Wan, K. Xiao, Q. Wan, D. Liu, and X. Liu, "rHDP: an aspect sharing-enhanced hierarchical topic model for multi-domain corpus," *ACM Transactions on Information Systems*, vol. 42, no. 3, pp. 1–31, May 2024, doi: 10.1145/3631352.

[17] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2018, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423.pdf

[18] M. Rey-Moreno, M. J. Sánchez-Franco, and M. de la S. Rey-Tienda, "Examining transaction-specific satisfaction and trust in Airbnb and hotels: an application of BERTopic and zero-shot text classification," *Tourism & Management Studies*, vol. 19, no. 2, pp. 21–37, Apr. 2023, doi: 10.18089/tms.2023.190202.

[19] A.-B. Bastidas-Manzano, J. Sánchez-Fernández, and L.-A. Casado-Aranda, "The past, present, and future of smart tourism destinations: a bibliometric analysis," *Journal of Hospitality & Tourism Research*, vol. 45, no. 3, pp. 529–552, Mar. 2021, doi: 10.1177/1096348020967062.

[20] M. Mariani and R. Baggio, "Big data and analytics in hospitality and tourism: a systematic literature review," *International Journal of Contemporary Hospitality Management*, vol. 34, no. 1, pp. 231–278, Jan. 2022, doi: 10.1108/IJCHM-03-2021-0301.

[21] Y. Guo, S. J. Barnes, and Q. Jia, "Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent Dirichlet allocation," *Tourism Management*, vol. 59, pp. 467–483, 2017, doi: 10.1016/j.tourman.2016.09.009.

[22] D. D'Amato *et al.*, "Green, circular, bio economy: a comparative analysis of sustainability avenues," *Journal of Cleaner Production*, vol. 168, pp. 716–734, Dec. 2017, doi: 10.1016/j.jclepro.2017.09.053.

[23] X. Xu and Y. Li, "The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: a text mining approach," *International Journal of Hospitality Management*, vol. 55, pp. 57–69, May 2016, doi: 10.1016/j.ijhm.2016.03.003.

[24] V. Taecharungroj and B. Mathayomchan, "Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand," *Tourism Management*, vol. 75, pp. 550–568, 2019, doi: 10.1016/j.tourman.2019.06.020.

[25] J.-W. Bi, Y. Liu, Z.-P. Fan, and J. Zhang, “Wisdom of crowds: conducting importance-performance analysis (IPA) through online reviews,” *Tourism Management*, vol. 70, pp. 460–478, Feb. 2019, doi: 10.1016/j.tourman.2018.09.010.
[26] A. C. Calheiros, S. Moro, and P. Rita, “Sentiment classification of consumer-generated online reviews using topic modeling,” *Journal of Hospitality Marketing & Management*, vol. 26, no. 7, pp. 675–693, Oct. 2017, doi: 10.1080/19368623.2017.1310075.
[27] A. García-Pablos, M. Cuadros, and G. Rigau, “W2VLDA: almost unsupervised system for aspect-based sentiment analysis,” *Expert Systems with Applications*, vol. 91, pp. 127–137, Jan. 2018, doi: 10.1016/j.eswa.2017.08.049.
[28] T. Brandt, J. Bendler, and D. Neumann, “Social media analytics and value creation in urban smart tourism ecosystems,” *Information & Management*, vol. 54, no. 6, pp. 703–713, Sep. 2017, doi: 10.1016/j.im.2017.01.004.
[29] S. (Sixue) Jia, “Motivation and satisfaction of Chinese and U.S. tourists in restaurants: a cross-cultural text mining of online reviews,” *Tourism Management*, vol. 78, p. 104071, Jun. 2020, doi: 10.1016/j.tourman.2019.104071.
[30] M. Nilashi, A. Fallahpour, K. Y. Wong, and F. Ghabban, “Customer satisfaction analysis and preference prediction in historic sites through electronic word of mouth,” *Neural Computing and Applications*, vol. 34, no. 16, pp. 13867–13881, Aug. 2022, doi: 10.1007/s00521-022-07186-5.
[31] C. K. H. Lee, “How guest-host interactions affect consumer experiences in the sharing economy: new evidence from a configurational analysis based on consumer reviews,” *Decision Support Systems*, vol. 152, p. 113634, Jan. 2022, doi: 10.1016/j.dss.2021.113634.
[32] X. Zhuo and W.-T. Wang, “‘Value for money?’ Exploring the consumer experience on shared accommodation platforms: evidence from online reviews in China,” *Journal of Hospitality and Tourism Technology*, vol. 13, no. 3, pp. 542–558, Jun. 2022, doi: 10.1108/JHTT-03-2021-0087.
[33] V. Ongsakul, T. Kajla, S. Raj, T. T. Khoa, and Z. U. Ahmed, “Changing tourists’ preferences in the hotel industry amid COVID-19 pandemic,” *Journal of Hospitality and Tourism Technology*, vol. 13, no. 2, pp. 295–313, Apr. 2022, doi: 10.1108/JHTT-07-2020-0179.
[34] C. Guan, Y.-C. Hung, and W. Liu, “Cultural differences in hospitality service evaluations: mining insights of user generated content,” *Electronic Markets*, vol. 32, no. 3, pp. 1061–1081, Sep. 2022, doi: 10.1007/s12525-022-00545-z.
[35] M. Venugopalan and D. Gupta, “An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis,” *Knowledge-Based Systems*, vol. 246, p. 108668, Jun. 2022, doi: 10.1016/j.knosys.2022.108668.
[36] E. Christodoulou, A. Gregoriades, M. Pampaka, and H. Herodotou, “Personality-informed restaurant recommendation,” in *Information Systems and Technologies. WorldCIST 2022. Lecture Notes in Networks and Systems*, vol. 468 LNNS, A. Rocha, H. Adeli, G. Dzemyda, and F. Moreira, Eds., Cham: Springer, 2022, pp. 13–21. doi: 10.1007/978-3-031-04826-5_2.
[37] B. Garner, C. Thornton, A. Luo Pawluk, R. Mora Cortez, W. Johnston, and C. Ayala, “Utilizing text-mining to explore consumer happiness within tourism destinations,” *Journal of Business Research*, vol. 139, pp. 1366–1377, Feb. 2022, doi: 10.1016/j.jbusres.2021.08.025.
[38] J. Liu, Y. Yu, F. Mehraliyev, S. Hu, and J. Chen, “What affects the online ratings of restaurant consumers: a research perspective on text-mining big data analysis,” *International Journal of Contemporary Hospitality Management*, vol. 34, no. 10, pp. 3607–3633, Aug. 2022, doi: 10.1108/IJCHM-06-2021-0749.
[39] Y. Song, K. Liu, L. Guo, Z. Yang, and M. Jin, “Does hotel customer satisfaction change during the COVID-19? A perspective from online reviews,” *Journal of Hospitality and Tourism Management*, vol. 51, pp. 132–138, Jun. 2022, doi: 10.1016/j.jhtm.2022.02.027.
[40] T. Kajla, S. Raj, S. Sharma, M. Joshi, and A. Kaur, “Key preferences of tourists during COVID-19 pandemic in luxury hotels: evidence from qualitative data,” *Tourism and Hospitality Research*, vol. 22, no. 4, pp. 473–487, Oct. 2022, doi: 10.1177/14673584211066742.
[41] Z. Shang and J. M. Luo, “Topic modeling for hiking trail online reviews: analysis of the Mutianyu Great Wall,” *Sustainability*, vol. 14, no. 6, p. 3246, Mar. 2022, doi: 10.3390/su14063246.
[42] X. Chen, J. Li, W. Han, and S. Liu, “Urban tourism destination image perception based on LDA integrating social network and emotion analysis: the example of Wuhan,” *Sustainability*, vol. 14, no. 1, p. 12, Dec. 2021, doi: 10.3390/su14010012.
[43] F. Otay Demir, Ş. Yavuz Görkem, and G. Rafferty, “An inquiry on the potential of computational literary techniques towards successful destination branding and literary tourism,” *Current Issues in Tourism*, vol. 25, no. 5, pp. 764–778, Mar. 2022, doi: 10.1080/13683500.2021.1887100.
[44] R. K. Mishra, J. A. A. Jothi, S. Urolagin, and K. Irani, “Knowledge-based topic retrieval for recommendations and tourism promotions,” *International Journal of Information Management Data Insights*, vol. 3, no. 1, p. 100145, Apr. 2023, doi: 10.1016/j.jjimei.2022.100145.
[45] A. Afaq, L. Gaur, and G. Singh, “A latent Dirichlet allocation technique for opinion mining of online reviews of global chain hotels,” in *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, IEEE, Apr. 2022, pp. 201–206. doi: 10.1109/ICIEM54221.2022.9853114.
[46] V. Puri, S. Mondal, S. Das, and V. G. Vrana, “Blockchain propels tourism industry—an attempt to explore topics and information in smart tourism management through text mining and machine learning,” *Informatics*, vol. 10, no. 1, p. 9, Jan. 2023, doi: 10.3390/informatics10010009.
[47] A. Alsayat, “Customer decision-making analysis based on big social data using machine learning: a case study of hotels in Mecca,” *Neural Computing and Applications*, vol. 35, no. 6, pp. 4701–4722, Feb. 2023, doi: 10.1007/s00521-022-07992-x.
[48] M. Nilashi *et al.*, “Revealing travellers’ satisfaction during COVID-19 outbreak: moderating role of service quality,” *Journal of Retailing and Consumer Services*, vol. 64, p. 102783, Jan. 2022, doi: 10.1016/j.jretconser.2021.102783.
[49] R. Nunkoo, A. Sharma, N. P. Rana, Y. K. Dwivedi, and V. A. Sunnassee, “Advancing sustainable development goals through interdisciplinarity in sustainable tourism research,” *Journal of Sustainable Tourism*, vol. 31, no. 3, pp. 735–759, Mar. 2023, doi: 10.1080/09669582.2021.2004416.
[50] M. P. Mehta, G. Kumar, and M. Ramkumar, “Customer expectations in the hotel industry during the COVID-19 pandemic: a global perspective using sentiment analysis,” *Tourism Recreation Research*, vol. 48, no. 1, pp. 110–127, Jan. 2023, doi: 10.1080/02508281.2021.1894692.
[51] N. Vaish, N. Goel, and G. Gupta, “Machine learning techniques for sentiment analysis of hotel reviews,” in *2022 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, Jan. 2022, pp. 01–07. doi: 10.1109/ICCCI54379.2022.9740876.
[52] Y. Luo, T. Tong, X. Zhang, Z. Yang, and L. Li, “Exploring destination image through online reviews: An augmented mining model using latent Dirichlet allocation combined with probabilistic hesitant fuzzy algorithm,” *Kybernetes*, vol. 52, no. 3, pp. 874–897, Mar. 2023, doi: 10.1108/K-07-2021-0584.

BIOGRAPHIES OF AUTHORS

Valentinus Roby Hananto is a lecturer at College of Information Science and Engineering, Ritsumeikan University, since 2024. After completing his doctoral degree in information science and engineering in 2022, he spent two years as a faculty member in the Department of Information Systems at Universitas Dinamika, Surabaya, Indonesia. His primary research theme is natural language processing. He is currently affiliated with the E-Society Laboratory in the Information Systems Science and Engineering (ISSE) Course, Ritsumeikan University. He can be contacted by email: hananto@fc.ritsumei.ac.jp.

Vivine Nurcahyawati is a Ph.D candidate at the Faculty of Computing, Universiti Malaysia Pahang. She has seven years of experience as a System Analyst and Database Administrator and 17 years of teaching at the Department of Information System, Universitas Dinamika, Surabaya, Indonesia. Her research areas include data mining, natural language processing, and software engineering. She can be contacted by email: vivine@dinamika.ac.id.

Tutut Wurijanto is a graduate of the Bachelor's program at STIKOM and holds a master's degree in information technology from the Institut Teknologi 10 Nopember Surabaya. Currently, he has 24 years of teaching experience at the Universitas Dinamika, specifically in the Bachelor's program in Information Systems. His teaching expertise lies in the fields of computer networking, software engineering, and data mining. He is currently conducting research in Data Mining and Software Engineering. He can be contacted by email: tutut@dinamika.ac.id.

Titik Lusiani is a lecturer at Dinamika University with a master's degree in informatics engineering. She holds a professional educator certification in the field of information systems since 2011 and also holds a database competency certification as Oracle Certified Professional (OCP). She has 28 years of teaching experience (since 1997). Her research interests are databases, expert systems, and software engineering. She can be contacted by email: lusiani@dinamika.ac.id.

Mate Kovacs received his Ph.D. in engineering from Ritsumeikan University, Japan, in 2021. He currently serves as an assistant professor at the same university. He is affiliated with the Digital Governance Systems Laboratory in the College of Information Science and Engineering. His research interests include natural language processing and deep learning for business intelligence and decision support. He can be contacted by email: kovacsm@fc.ritsumei.ac.jp.