

Review Articles

UDC: 338.486.3:640.45]:303.425

005.521:334.7

DOI: 10.5937/menhottur2500010G

Received: 9 May 2025

Revised: 10 June 2025

Accepted: 20 July 2025

Published online: 13 August 2025

Topic modeling in hospitality and tourism research: Application areas, business insights, and managerial implications

Olivera Grljević^{1*}

¹ University of Novi Sad, Faculty of Economics in Subotica, Serbia

Abstract

Purpose – Topic modeling (TM) explores customer experience and behaviors from large volumes of textual data, such as online reviews uncovering (dis)satisfaction cues often overlooked by hospitality managers. Despite its potential, TM application in hospitality research is limited compared to other social science methods. This paper aims to investigate the scope of TM research in the hospitality domain and contribute to the understanding of the areas where it can be effectively applied, the purposes it can serve, and the types of problems it can address. **Methodology** – The research methodology is rooted in the systematic literature review – 40 relevant papers were collected and analysed to identify the areas of hospitality where TM is mostly applied, business insights derived from TM application, and commonly utilised TM approaches. **Findings** – TM research in hospitality is conducted in five research areas: accommodation and lodging, food and beverages, attractions and events, nature-based tourism, and travel services. Researchers apply TM to gain nine different business insights, such as dissatisfaction drivers, segment-based preferences, sentiments, preference changes over time, service quality perception, or underexplored areas. **Implications** – TM-based research provides actionable recommendations for the enhancement of managerial practices within the hospitality industry, such as promotion and destination management, service improvements, and reduction of overtourism.

Keywords: topic modeling, natural language processing, hospitality, user-generated content

JEL classification: Z3, C55, C49

Modelovanje tema u istraživanjima iz oblasti ugostiteljstva i turizma: Oblasti primene, poslovni uvidi i implikacije

Sažetak

Svrha – Modelovanje tema (*topic modeling* – TM) istražuje iskustvo i ponašanje potrošača analizom velikih količina tekstualnih podataka, poput onlajn recenzija. Menadžeri u ugostiteljstvu često previđaju ovako otkrivene znakove (ne)zadovoljstva. Uprkos potencijalu, primena TM u ugostiteljstvu je ograničena u poređenju sa drugim metodama društvenih nauka. Rad ima za cilj da ispita TM istraživanja u oblasti ugostiteljstva i doprinese

* Corresponding author: olivera.grljevic@ef.uns.ac.rs



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

razumevanju oblasti u kojima se TM može efikasno primeniti i poslovne probleme koje može adresirati. **Metodologija** – Sistematskim pregledom literature identifikovano je 40 relevantnih radova koji su analizirani kako bi se prepoznale oblasti ugostiteljstva u kojima se TM najviše primenjuje, poslovni uvidi izvedeni primenom TM i najčešće korišćeni pristupi. **Rezultati** – Istraživači primenjuju TM u pet istraživačkih oblasti: smeštaj, hrana i piće, atrakcije i događaji, turizam zasnovan na prirodnim resursima i putničke usluge. Devet različitih vrsta poslovnih uvida izvodi se primenom TM, poput faktora koji utiču na nezadovoljstvo, preferencije po segmentima, sentiment, promene u preferencijama tokom vremena, percepcija o kvalitetu usluga ili nedovoljno istražena područja. **Implikacije** – TM istraživanja pružaju preporuke za unapređenje ugostiteljsko-menadžerske prakse, poput promocije, unapređenja usluga ili smanjenja prekomernog turizma.

Ključne reči: modelovanje tema, obrada prirodnog jezika, ugostiteljstvo, korisnički generisani sadržaj

JEL klasifikacija: Z3, C55, C49

1. Introduction

The Internet and social media have transformed the way people seek and exchange hospitality-related information. Electronic word-of-mouth advertising has replaced the traditional form (Gruen et al., 2006; Liu et al., 2024), allowing individuals to consult online platforms for insights before making purchases or travel-related decisions. After consumption, people share personal experiences in the form of online reviews and ratings. Social media has thus evolved into a valuable channel for capturing the voice of the customers. It offers hospitality managers and marketers valuable input for strategic decisions, improvement of customer engagement, and service offerings. However, traditional analytical approaches cannot handle the scale and complexity of its data (Laureate et al., 2023). Advanced computational resources are required for dealing high data volume and the prevalence of unstructured textual data on social media, estimated at over 80% of the data (Anandarajan et al., 2019), requires the application of artificial intelligence (AI) for effective processing, such as natural language processing (NLP) techniques. In response, techniques and approaches for handling and analysing social media data continue to evolve (Grljević & Marić, 2024). With advancement of AI it becomes imperative for studies to transition toward examining actual customer experiences and measure their real behaviours (Gursoy & Cai, 2025), contrary to traditional use of qualitative methods to gain insights about customer perception (Gursoy & Cai, 2025) or use of hypothetical scenarios to simulate real-life conditions (Law et al., 2024; Xu et al., 2023). To support this shift, this study focuses on topic modeling (TM), a method that enables researchers to analyse actual customer-generated content and derive insights grounded in authentic behavioural data.

TM is an NLP technique applied to a collection of documents to uncover hidden thematic structures within data collection, referred to as corpus, aiding automated information summarization and data comprehension (Maier et al., 2018). Documents can refer to any text, such as one piece of text published on social media platform. TM operates under the assumption that documents are expressed using a limited number of core concepts, called topics. The objectives of TM are to identify topics, to indicate connections between topics, and to track their evolution over time. It is beneficial for monitoring trends, sentiments, rumours, and the factors influencing people's consumption of services or products (Grljević & Marić, 2024). In hospitality, TM is particularly valuable as it often reveals insights into customer experience that are otherwise missed or neglected by marketers (Li et al., 2018; Shafqat & Byun, 2020). It allows for identification of factors contributing to customer satisfaction and root causes of dissatisfaction (Grljević et al., 2025) enabling hospitality

managers to use the obtained insights for improvement of customer experience, satisfaction, and loyalty (Nguyen & Ho, 2023).

TM can be used as a stand-alone technique for exploratory or descriptive analysis (Banks et al., 2018; Gao et al., 2022; Nguyen & Ho, 2023), or as part of a more comprehensive AI-based systems, such as a recommendation system (Shafqat & Byun, 2020). However, Papilloud and Hinneburg (2018) found that literature on TM application in hospitality is limited compared to other methods applied in social research. This is due to the present gap in knowledge and skills required to effectively employ it in hospitality-related research. The gap encompasses areas like programming, statistics, and the fundamentals of mathematical modeling (Papilloud & Hinneburg, 2018), indicating the need for stronger interdisciplinary collaboration among researchers. In order for this valuable technique to gain broader practical adoption, a more mature and systematic body of research is needed that will provide researchers understanding of the areas where TM can be effectively applied, the purposes it can serve, and the types of problems it can address. To the best of our knowledge, no previous studies has specifically focused on the applicative use of topic modeling in the hospitality sector. This motivated the research presented in the current study, aiming to characterise TM research in hospitality sector by assessing the scope of TM applications in the hospitality domain and its potential effect on the sector. This distinguishes our work from previous literature reviews that primarily focus on the broader fields, such as artificial intelligence (Gursoy & Cai, 2025; Kim et al., 2025; Law et al., 2024) or machine learning (Núñez et al., 2024) and their applications in hospitality and tourism, rather than on detailed examination of single technique, such as topic modeling and how it is employed as a tool for business knowledge discovery.

The following research questions are defined.

RQ1: What are the areas of hospitality where TM is most applied?

RQ2: What business insights can be derived from TM application in hospitality domain?

RQ3: Which TM approaches (algorithms) are mostly utilized in the hospitality research?

The paper is structured as follows. The second section presents research methodology. Results and discussion are presented in the third section and structured to answer each of the proposed research questions. Practical implications are presented in the fourth section, while concluding remarks are presented in the final section of the paper.

2. Methodology

The research methodology is rooted in a systematic literature review (SLR) (Kitchenham & Charters, 2007). SLR introduces research rigor and allows for obtaining insights that are aligned with the proposed research questions. In this paper, SLR is utilized to comprehensively analyze and document the scientific development of TM in hospitality research. Kitchenham and Charters (2007) suggest three stages of SLR, followed in this paper.

Phase 1 refers to planning the SLR process, where the need, scope, and objectives of SLR are examined. This includes decisions on the academic databases that will be used for literature search and the criteria for inclusion and exclusion of papers from the SLR. This phase results in a search strategy protocol.

Phase 2 is dedicated to conducting an SLR during which the literature is searched by the strategy set in the previous phase. Defined inclusion and exclusion criteria filter studies to retain only the most relevant ones, which are further assessed for quality.

Phase 3 refers to reporting on the results, which are presented in Section 3 of this paper. Conclusions are drawn from the results. Both results and conclusions are aligned with the proposed research questions.

2.1. Selection criteria

Articles were collected through a search of the Web of Science (WoS) and Scopus literature databases. These databases are selected as they comply with high-quality criteria and are cross-linked with many other databases. According to [Kirilenko et al. \(2021\)](#), TM-based research gained momentum in 2019. Motivated by their findings, the search was restricted to peer-reviewed articles published in the five years span (2019-2023) in scientific journals, while the paper collection was conducted in January 2024. The language is restricted to English as the predominant language in scientific literature.

To identify relevant studies, search keywords are set as a combination of terms (including synonyms, alternative spellings, and related terms) with Boolean operators. “Topic modeling” or “topic modelling” is the main concept. The application domains are combined with the main concept using the OR operator: “tourism” OR “destination marketing” OR “tourist perceptions” OR “hospitality” OR “attractions” OR “tourists’ preferences” OR “customer satisfaction” OR “destination management”.

The search is narrowed to the categories that are in the research focus. In WoS, the selected categories are: Hospitality Leisure Sport Tourism; Management; Business; Environmental Science; Environmental Studies; and Green Sustainable Science Technology. These six categories are also the predominant categories in which papers are published. Within selected categories, the search was further narrowed by filtering out subcategories not correlated with the hospitality industry, such as engineering, marine freshwater biology, toxicology, and geography. In Scopus, the selected categories included: Business, Management and Accounting; Social Sciences; Environmental Science; Economics, Econometrics and Finance; Decision Science; and Computer Science, given that this category was in the top three categories according to the number of published articles. The search resulted in 502 publications from WoS and 213 from Scopus.

2.2. Paper screening

SLR sought to identify articles where TM was employed to investigate some phenomena in the hospitality industry. A screening process was conducted to identify and exclude research that did not meet this criterion. Analysis of the articles not highlighting TM in the title or abstract indicated that researchers did not employ TM. These articles were excluded. A total of 81 articles from WoS and 110 articles from Scopus remained. An additional 54 WoS and 83 Scopus articles were removed following the inclusion and exclusion criteria presented in Table 1.

Table 1: Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Publications in the scope of hospitality	Publications out of hospitality scope
TM is a core method	TM is not a core method
TM is utilized to investigate phenomena in hospitality industry	TM is applied to categorize hospitality literature
The main data type is short text	The main data type is not short text
Non-COVID studies	COVID-related studies
Publications available through institutional subscriptions	Publications unavailable through institutional subscriptions

Source: Author's research

Upon analysing articles and applying these criteria, we identified: 1) Thirty-three articles are characterized as out-of-scope and irrelevant (18 WoS and 15 Scopus) as they focus on various unrelated topics, such as e-government, adaptive clothing customers, and mobile banking. 2) Fifteen articles (2 WoS and 13 Scopus) did not use TM as a core method. As TM represents one of the steps in their overall methodological framework, information and a detailed description of the methodology related to TM are omitted in these papers, as well as a more extensive interpretation of the results of TM. 3) Twenty-three articles (8 WoS and 15 Scopus) that used TM for literature review were excluded as their main focus is bibliometric or scientometric analysis rather than investigating specific hospitality phenomena. 4) Twenty-nine articles (11 WoS and 18 Scopus), although within the research scope, dealt with non-textual data types and were subsequently removed. 5) Twenty-five COVID-related studies (11 WoS and 14 Scopus) were excluded as they address atypical business conditions caused by the pandemic. The insights derived from these studies cannot be generalized to the typical conditions of the hospitality industry. 6) Twelve articles (4 WoS and 8 Scopus) could not be retrieved through institutional subscriptions and were not considered in the SLR.

The resulting papers overlapped in 14 articles, which were excluded. After the screening, 40 articles remained for the SLR, 27 WoS and 13 Scopus articles.

3. Results and discussion

Table 2 presents the resulting collection of studies with respective IDs used in this paper.

Table 2: Studies on TM application in hospitality domain

Publication reference	ID
(Sánchez-Franco & Aramendia-Muneta, 2023)	P1
(Nguyen & Ho, 2023)	P2
(Srinivas & Ramachandiran, 2024)	P3
(Gregoriades et al., 2023)	P4
(Zolfaghari & Choi, 2023)	P5
(Taecharungroj, 2023)	P6
(Shang et al., 2022)	P7
(Kwon et al., 2022)	P8
(Egger & Yu, 2022)	P9

(Shang & Luo, 2022)	P10
(Gao et al., 2022)	P11
(Wu et al., 2022)	P12
(Tang et al., 2022)	P13
(Garner et al., 2022)	P14
(Viñán-Ludeña & de Campos, 2022)	P15
(Wang et al., 2021)	P16
(Mirzaalian & Halpenny, 2021)	P17
(Twil et al., 2021)	P18
(Marcolin et al., 2021)	P19
(Kirilenko et al., 2021)	P20
(Luo et al., 2021)	P21
(Vargas-Calderón et al., 2021)	P22
(Sim et al., 2021)	P23
(Janssens et al., 2021)	P24
(Celuch, 2021)	P25
(Han & Yang, 2021)	P26
(Kar et al., 2021)	P27
(Park et al., 2020)	P28
(Luo et al., 2020)	P29
(Ding et al., 2020)	P30
(Sutherland & Kiatkawsin, 2020)	P31
(Shafqat & Byun, 2020)	P32
(Wen, et al., 2020)	P33
(Kwon et al., 2020)	P34
(Celata et al., 2020)	P35
(Aggarwal & Gour, 2020)	P36
(Hu et al., 2019)	P37
(Kim et al., 2019)	P38
(Zhang, 2019)	P39
(Sanchez-Franco et al., 2019)	P40

Source: Author's research

To answer the research questions authors extracted and analysed data from the collected studies on: a) social media platform used as a data source in the research, b) data focus indicating the content or the subject of collected data, and c) TM approach – an algorithm utilized to extract topics from the data collection. The results of the analysis are presented in the following subsections.

3.1. Corpus origins and data utilization in TM-based hospitality research

Table 3 presents information on social media and data focus with respective study references.

Table 3: Characteristics of TM-based hospitality studies

Social media	No. of studies	Data focus	References
TripAdvisor	6	accommodation reviews	P1, P12, P4, P11, P37, P19
	4	reviews of tourist localities or attractions	P18, P20, P29, P36
	7	reviews of natural sites	P5, P6, P7, P10, P13, P17, P32
Airbnb & Insideairbnb	8	reviews of private accommodation, lodging or short-term rentals	P1, P11, P24, P26, P30, P31, P35, P39
Agoda	1	reviews of hotel products and services	P2
	1	reviews of accommodation	P23
Booking	2	hotel reviews	P4, P22
Yelp	1	review of airline services	P3
	1	reviews of tourist attractions	P14
	1	hotel reviews	P40
	3	restaurant reviews	P8, P28, P34
AllergyEats.com	1	restaurant reviews	P33
Instagram	1	dark tourism	P9
Google reviews	1	reviews of natural sites	P32
Sina Weibo	1	reviews of natural sites	P16
DianPing	1	reviews of natural sites	P10
Ctrip	2	reviews of natural sites	P10, P21
Qunar and Baidu Travel	2	reviews of natural sites	P10, P21
Tongcheng Travel	1	reviews of natural sites	P21
Trustpilot	1	reviews of event services	P25
Twitter	2	reviews of tourist localities or attractions	P15
		service experience	P27
Skytrax	1	airline service reviews	P3

Source: Author's research

TM-based hospitality research utilizes data from 16 different social media platforms, with TripAdvisor being the most prominent data source (42.5%), followed by Airbnb and insideairbnb.com (20%), and Yelp (15%). Predominantly, authors retrieved data from one

data source, while in 17.5%, more than one and a maximum of four different data sources are used (two: P1, P3, P4, P11, P32; three: P21; four: P10). The additional data source is introduced to achieve some specific research goals related to differences in language characteristics (P1, P32) or cultural variations (P4). The authors of one article do not report on data source P38.

Except for Instagram, Twitter, and Sina Weibo – a Chinese microblogging website, all social media are review platforms enabling users to freely express their opinion or reflect on personal customer experience. Therefore, the main data type used in hospitality research is online reviews. Authors also use narrative accompanying Instagram photos (P9) or microposts (P15, P16, and P27). These data types imply specific characteristics, such as shortness and colloquial writing style without respect to grammar or spelling, influencing the effectiveness of applied TM algorithms (Tang et al., 2014).

Data focus implies focused data collection on: accommodation, restaurants, localities, attractions, natural resources, or hospitality-related services.

3.2. Application areas of TM in hospitality research

The main data focus, presented in Table 3, indicates that TM-based hospitality research can be grouped into five general research areas: *Accommodation and lodging* (40% of studies), *Food and beverages* (10%), *Attractions and events* (22.5%), *Nature-based tourism* (22.5%), and *Travel services* (5%). Table 4 provides an overview of studies based on the identified research areas. These studies explore customers' preferences or perceptions within the respective hospitality topic. Preferences refer to how much customers appreciate, value, and notice various features of hospitality-related services or products. Customers' perception is associated with the sensory experience of the service or products.

Table 4: Areas of TM applications in hospitality research

Research areas	No. of studies	References
Accommodation and lodging	16	P1, P2, P4, P11, P12, P19, P22, P23, P24, P26, P30, P31, P35, P37, P39, P40
Food and beverages	4	P8, P28, P33, P34
Attractions and events	9	P9, P14, P15, P18, P20, P25, P29, P36, P38
Nature-based tourism	9	P5, P6, P7, P10, P13, P16, P17, P21, P32
Travel services	2	P3, P27

Source: Author's research

Accommodation and lodging-related research is closely associated with the analysis of tourists' preferences. Research focuses on the evaluation of customer experience by examining factors influencing (dis)satisfaction of customers (P1, P2, P11, P12, P19, P31, P37) or by evaluating the service quality of hotels or alternative accommodation and lodgings (P22, P30, P39). Some of the research explores variations of identified factors across hotel ratings (P11, P37), price levels (P11), room type (P26), customers' scores (P24, P39), seasonal changes (P30), or topic relevance across geographical segments – the spatial analysis (P35). Factors influencing customer satisfaction are explored in terms of building loyalty (P40) and revisit intention (P4). Authors of the P23 study differ in perspective as they measure the influence of listing descriptions on listing demand.

Research related to *food and beverages* naturally builds upon analysis of customers' perception and their sensory experience. Authors explore customers' perception regarding specific phenomena within the hospitality industry, such as perception of green restaurant practices (P28), luxury consumption in restaurants (P8), or restaurants accommodating allergen-free requests (P33). By researching customer perception, the authors aim to identify key factors influencing customer (dis)satisfaction. Only one research tracks changes in customer perception over time (P28).

Attractions and events-related studies adopted a more exploratory approach. Authors use TM to uncover general or topics most commonly discussed among visitors of various tourist attractions, such as cultural heritage and historic sites (P9, P18, P20, P38), entertainment and leisure sites (P29), or museums (P20). By examining topics and extracting new insights and knowledge from topics, researchers strived to contribute to and enhance managerial practices in the hospitality industry in the following manner: a) *Monitoring changes in topics over time*, which offers practical insights for destination marketing and management; e.g., P38 identified the visitors of Jeju Island UNESCO heritage sites became less satisfied with visiting heritage sites and began seeking more adventurous experiences in these areas over time; b) *Customer satisfaction monitoring*, which provides insights valuable for effective customer dissatisfaction management and helps improve overall customer experience (P18, P20, P36); furthermore, it is closely related to c) *Identification of sentiments* associated with tourists' experience (P18, P14, P15). Authors in P14 study explore worldwide attractions to understand short-term happiness and factors influencing travel satisfaction, providing insights into memorable experiences, while in P15, authors analyzed Granada's tourism, considering places, events, restaurants, and hotels. Their study incorporates a seasonal perspective and offers insights for practitioners and travellers by providing an overview of popular places, major attractions, and valuable information for managing tourism products and targeting influential users for joint promotional activities (P15). One study addresses the events industry by investigating customer sentiments related to the experience of purchasing tickets from a third-party platform to improve overall customer experience (P25).

Research related to *nature-based tourism* explores topics regarding natural reserves and national parks (P5, P16, P17, P21), recreational and outdoor activities (P6, P7, P10), and less explored destinations, such as Jeju Islands' under-emphasized tourist spots (P32) or glacier tourism destinations (P13). These studies focus on the exploration of visitors' preferences and perceptions, with the outcomes of TM carrying managerial implications. Authors of P32 explored possibilities of enhancing recommendations of less explored destinations, while authors of P17 analysed possibilities of application of TM and sentiment analysis for destination marketing and evaluation of destination loyalty. Three studies particularly differ in their perspective of tourists' perception as authors of P7 explore seasonal changes in attributes affecting tourists' perception, P16 analyse differences in perception among on-site and after-trip groups of tourist reviews, while in P10 authors analyse cultural differences in tourist perception through most-talked about topics in reviews of national and international tourists.

Studies related to travel services contribute to improved understanding of customer preferences. Authors of the P3 aim to identify key factors affecting passenger (dis)satisfaction in the airline industry according to each of the service aspects. In P27, authors recognize the importance and impact of factors influencing customer service experiences across different zones of India. The authors highlighted the need for adopting approaches specific to locations or zones when managing customer services or tailoring customer service strategies.

3.3. Business insights supported by TM in hospitality research

Analysis of studies within identified research areas also provides an understanding of the business insights that can be obtained through TM, which can further be utilized for addressing business problems. These insights could be grouped into nine categories, as presented in Table 5.

Table 5: Summary of TM-supported business and research topics

Analytical topic	No. of studies (%)	Research area	References
Identification of factors influencing customer satisfaction and dissatisfaction	16 (40%)	Accommodation and lodging	P1, P2, P4, P11, P12, P19, P31, P37, P40
		Food and beverages	P28, P8, P33
		Attractions and events	P18, P20, P36
		Travel services	P3
Segment-based analysis of perceptions and preferences	8 (20%)	Accommodation and lodging	P11, P24, P26, P35, P37, P39
		Nature-based tourism	P10, P16
Sentiment and emotion analysis of visitor experiences	5 (12.5%)	Attractions and events	P14, P15, P18, P25
		Nature-based tourism	P17
Monitoring changes in customer preferences over time	5 (12.5%)	Accommodation and lodging	P30
		Food and beverages	P28
		Attractions and events	P38
		Nature-based tourism	P7, P16
Evaluation of service quality and performance	5 (12.5%)	Accommodation and lodging	P22, P30, P39
		Travel services	P3, P27
Enhancement of destination marketing and product positioning	4 (10%)	Nature-based tourism	P17, P32
		Attractions and events	P15, P38
Identification of hidden or less explored areas and attractions	2 (5%)	Nature-based tourism	P13, P32
Optimization of digital content	1 (2.5%)	Accommodation and lodging	P23
Support for joint promotional and influencer strategies	1 (2.5%)	Attractions and events	P15

Source: Author's research

Identification of factors influencing customer satisfaction and dissatisfaction is the prevailing analytical topic (40% of studies), followed by segment-based analysis of perceptions and preferences (20%). Sentiment and emotion analysis of visitors experiences, monitoring changes in customer preferences over time and evaluation of service quality and performance are equally present in 12.5% of the studies each. Most of the studies are based on one

analytical topics, while 27.5% address two (P3, P11, P16, P17, P18, P28, P30, P32, P37, P38, P39) and only one study addresses three analytical topics (P15).

The results indicate that certain categories are present across wider range of hospitality-related research areas. Namely, identification of influential (dis)satisfaction factors or tracking changes in customer preferences over time characterize four different research areas, while others are more specific, such as identification of hidden or less explored localities, optimization of digital content, and support of joint promotional activities that are specific to one research area only, i.e. *nature-based tourism*, *accommodation and lodging*, and *attractions and events*, respectively. These findings could direct researchers to broaden the research scope and explore possibilities of addressing other problems in the respective research area through TM application.

3.4. TM algorithms in hospitality research

Authors use nine different TM approaches, i.e., algorithms, in the hospitality research, as presented in Table 6. Latent Dirichlet Allocation (LDA) is used in 67.5% of the studies. Structural Topic Model (STM) is the second most commonly employed approach in hospitality research. It is used in 22.5% of studies, while Non-negative Matrix Factorisation (NMF) is used in 5%. Other approaches, i.e. Contextual Topic Model (CxTM), Dynamic Topic Model (DTM), Latent Semantic Analysis (LSA), Correlation Explanation (CorEx), Correlated Topic Model (CTM), and Probabilistic Latent Semantic Analysis (pLSA), are each used in a single study. These findings are in line with [Laureate et al.'s \(2023\)](#) findings about LDA being the prominent approach across all scientific research areas in which it has been applied so far. They also found that the LDA application is not always justified or optimal. LDA has proven effectiveness on longer documents, as opposed to short texts, such as online reviews ([Laureate et al., 2023](#); [Mazarura & De Waal, 2016](#); [Tang et al., 2014](#); [Yan et al., 2013](#); [Zou & Song, 2016](#)). This indicates the need for more research on assessing the effectiveness of the LDA algorithm in the hospitality domain.

Table 6: TM approaches according to research areas

TM approach	No. of studies	Research area	References
LDA	27	Accommodation and lodging	P2, P12, P22, P23, P24, P31, P39, P40
		Attractions and events	P9, P15, P18, P20, P25, P29, P36, P38
		Nature-based tourism	P5, P6, P7, P10, P13, P16, P17, P21, P32
		Travel services	P3, P27
STM	9	Accommodation and lodging	P11, P4, P26, P30, P37
		Food and beverages	P8, P28, P33, P34
NMF	2	Accommodation and lodging	P35
		Attractions and events	P9
CxTM	1	Accommodation and lodging	P1
DTM	1	Accommodation and lodging	P2
LSA	1	Accommodation and lodging	P19
CorEx	1	Attractions and events	P9
CTM	1	Attractions and events	P14
pLSA	1	Travel services	P3

Source: Author's research

LDA is applied in all research areas in hospitality, except in *food and beverages*, which is completely based on the application of STM. STM is a more appropriate approach for studying relations between metadata (e.g., price levels, ratings) and topics, and to estimate how they affect the generation of text (Roberts et al., 2014). Studies in *nature-based tourism* research are also completely characterized by a singular TM approach – the LDA. In other research areas, authors applied various TM approaches.

4. Practical implications

Practical implications are discussed in the context of business insights categories presented in Table 5. Identification of factors contributing to customer dissatisfaction (P1, P2, P3, P4, P8, P11, P12, P18, P19, P20, P28, P31, P33, P36, P37, P40) enables hospitality managers to work on improving services to better adhere to customer needs, leading to improved customer experience and loyalty (Grljević et al., 2025), while strengthening revisiting intentions. TM is often combined with sentiment analysis, which broadens analysis of influential factors with customers' sentiments and emotions (P14, P15, P17, P18, and P25). The results could assist hospitality managers in efficiently addressing the emotional responses of visitors or service users, addressing safety concerns, and enhancing trust and perception. These analyses are often segment-based, helping to reveal perceptions and preferences within defined segments, such as price ranges (P11), grades (P11 and P37), geo-locations (P35), cultural origins (P10 and P16), or seasons (P30). Segment-based analysis improves segmentation strategies and enables personalization of offerings. Customer preferences could be monitored over time (P7, P16, P28, P30, and P38) to track changes in preferences and adjust marketing activities or services based on evolving customer needs or expectations. If TM is used for evaluation of service quality (P3, P22, P27, P30, and P39), it may assist in benchmarking, identification of service gaps, and direct improvements accordingly.

Insights gained through TM may also be used for campaign adjustments to promote destinations, optimize marketing, or identify unique selling points that will be further highlighted in promotions. In this way, TM contributes to the enhancement of destination marketing or product positioning (P15, P17, P32, and P38). TM has proven its effectiveness in the identification of hidden or less explored areas and attractions (P13 and P32). By revealing insights in this respect, TM results may contribute to the encouragement of tourist dispersion, reduction of overtourism, and promotion of alternative destinations. Concerning promotional efforts, TM may assist in targeting influential users for marketing and brand positioning (P15), as well as for optimization of digital content (P23), such as listing descriptions. Improved content strategy may positively affect bookings and engagement.

5. Conclusion

TM represents an effective way to understand large volumes of data about customer experience in the hospitality domain. The research based on TM analysis, as results suggest, is grouped into five distinct application areas: accommodation and lodging, food and beverages, attractions and events, nature-based tourism, and travel services. Insights gained through TM span a wide range, from identification of (dis)satisfaction drivers, sentiments, perceptions and preferences among various segments, to changes over time, service quality evaluation, or identification of hidden or less explored areas and attractions. These insights have a direct impact on managerial activities in the hospitality industry and may enhance them, such as to inform strategies for improvement of customer experience and loyalty, help

to identify service gaps and direct improvements, optimize marketing according to customers' changing preferences, manage overtourism, or conduct benchmarking.

Although the presented study offers valuable insights into the characteristics of TM research in the hospitality domain, limitations should be acknowledged. Current research is built upon the literature collected from two databases, WoS and Scopus, while other databases were not consulted. The decision was motivated by the fact that WoS and Scopus databases are cross-linked with other databases and they ensure high quality criteria through strict peer-reviewing process. A time span refreshment is needed to encompass recent studies and evaluate the impact of large language models on TM applications in hospitality research. Future work may explore the transferability of TM approaches between hospitality subfields and evaluate the potential for transferring research frameworks across different TM application areas, to assess possibilities to generate similar types of business insights in all application areas.

CRedit author statement

The author is responsible for all aspects of the research and manuscript preparation.

Declaration of generative AI in the writing process

During the preparation of the paper, the author used Grammarly and ChatGPT in a complementary manner, solely to improve the linguistic quality of the manuscript, the clarity and the fluency of the English language. In particular, when Grammarly suggestions were challenging to interpret or resolve independently, the author used ChatGPT to better understand and refine those specific language issues. After using this service, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Acknowledgment

This paper is part of a broader research initiative on TM application in hospitality research. Two distinct studies have been conducted drawing from a shared body of literature identified through a SLR: the present paper, and a separate study focused on the development of a conceptual model for short-text TM and practical guidelines for researchers and practitioners. Sincere thanks to Assistant Professor Nebojša Taušan, University of Novi Sad, Faculty of Economics in Subotica, for his valuable contribution to the SLR process.

Conflict of interest

The author declares no conflict of interest.

References

1. Aggarwal, S., & Gour, A. (2020). Peeking inside the minds of tourists using a novel web analytics approach. *Journal of Hospitality and Tourism Management*, 45, 580–591. <https://doi.org/10.1016/j.jhtm.2020.10.009>
2. Anandarajan, M., Hill, C., & Nolan, T. (2019). *Practical text analytics: Maximizing the value of text data*. Springer Cham. <https://doi.org/10.1007/978-3-319-95663-3>

3. Banks, G. C., Woznyj, H. M., Wesslen, R. S., & Ross, R. L. (2018). A review of best practice recommendations for text analysis in R (and a user-friendly app). *Journal of Business and Psychology*, 33, 445–459. <https://doi.org/10.1007/s10869-017-9528-3>
4. Celata, F., Capineri, C., & Romano, A. (2020). A room with a (re)view. Short-term rentals, digital reputation and the uneven spatiality of platform-mediated tourism. *Geoforum*, 112, 129–138. <https://doi.org/10.1016/j.geoforum.2020.04.007>
5. Celuch, K. (2021). Customers' experience of purchasing event tickets: Mining online reviews based on topic modeling and sentiment analysis. *International Journal of Event and Festival Management*, 12(1), 36–50. <https://doi.org/10.1108/IJEFM-06-2020-0034>
6. Ding, K., Choo, W. C., Ng, K. Y., & Ng, S. I. (2020). Employing structural topic modelling to explore perceived service quality attributes in Airbnb accommodation. *International Journal of Hospitality Management*, 91, 102676. <https://doi.org/10.1016/j.ijhm.2020.102676>
7. Egger, R., & Yu, J. (2022). Identifying hidden semantic structures in Instagram data: A topic modelling comparison. *Tourism Review*, 77(4), 1234–1246. <https://doi.org/10.1108/TR-05-2021-0244>
8. Gao, B., Zhu, M., Liu, S., & Jiang, M. (2022). Different voices between Airbnb and hotel customers: An integrated analysis of online reviews using structural topic model. *Journal of Hospitality and Tourism Management*, 51, 119–131. <https://doi.org/10.1016/j.jhtm.2022.03.004>
9. Garner, B., Thornton, C., Pawluk, A. L., Cortez, R. M., Johnston, W., & Ayala, C. (2022). Utilizing text-mining to explore consumer happiness within tourism destinations. *Journal of Business Research*, 139, 1366–1377. <https://doi.org/10.1016/j.jbusres.2021.08.025>
10. Gregoriades, A., Pampaka, M., Herodotou, H., & Christodoulou, E. (2023). Explaining tourist revisit intention using natural language processing and classification techniques. *Journal of Big Data*, 10(1), 1–31. <https://doi.org/10.1186/s40537-023-00740-5>
11. Grljević, O., & Marić, M. (2024). A comprehensive analysis of online reviews in the Srem region through topic modeling. In V. Bevanda, & S. Štetić (Eds.), *8th International Thematic Monograph: Modern Management Tools and Economy of Tourism Sector in Present Era* (pp. 291-311). Belgrade, Serbia: Association of Economists and Managers of the Balkans in cooperation with the Faculty of Tourism and Hospitality, Ohrid, North Macedonia. <https://doi.org/10.31410/tmt.2023-2024.291>
12. Grljević, O., Marić, M., & Božić, R. (2025). Exploring mobile application user experience through topic modeling. *Sustainability*, 17(3), 1109. <https://doi.org/10.3390/su17031109>
13. Gruen, T. W., Osmonbekov, T., & Czaplewski, A. J. (2006). eWOM: The impact of customer-to-customer online know-how exchange on customer value and loyalty. *Journal of Business Research*, 59(4), 449–456. <https://doi.org/10.1016/j.jbusres.2005.10.004>
14. Gursoy, D., & Cai, R. (2025). Artificial intelligence: An overview of research trends and future directions. *International Journal of Contemporary Hospitality Management*, 37(1), 1–17. <https://doi.org/10.1108/IJCHM-03-2024-0322>
15. Han, C., & Yang, M. (2021). Revealing Airbnb user concerns on different room types. *Annals of Tourism Research*, 89, 103081. <https://doi.org/10.1016/j.annals.2020.103081>
16. Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management*, 72, 417–426. <https://doi.org/10.1016/j.tourman.2019.01.002>
17. Janssens, B., Bogaert, M., & Van den Poel, D. (2021). Evaluating the influence of Airbnb listings' descriptions on demand. *International Journal of Hospitality Management*, 99, 103071. <https://doi.org/10.1016/j.ijhm.2021.103071>

18. Kar, A. K., Kumar, S., & Ilavarasan, P. V. (2021). Modelling the service experience encounters using user-generated content: A text mining approach. *Global Journal of Flexible Systems Management*, 22, 267–288. <https://doi.org/10.1007/s40171-021-00279-5>
19. Kim, H., So, K. K. F., Shin, S., & Li, J. (2025). Artificial intelligence in hospitality and tourism: Insights from industry practices, research literature, and expert opinions. *Journal of Hospitality & Tourism Research*, 49(2), 366–385. <https://doi.org/10.1177/10963480241229235>
20. Kim, K., Park, O., Barr, J., & Yun, H. (2019). Tourists' shifting perceptions of UNESCO heritage sites: Lessons from Jeju Island-South Korea. *Tourism Review*, 74(1), 20–29. <https://doi.org/10.1108/TR-09-2017-0140>
21. Kirilenko, A. P., Stepchenkova, S. O., & Dai, X. (2021). Automated topic modeling of tourist reviews: Does the Anna Karenina principle apply? *Tourism Management*, 83, 104241. <https://doi.org/10.1016/j.tourman.2020.104241>
22. Kitchenham, B., & Charters, S. M. (2007). *Guidelines for performing systematic literature reviews in software engineering*. Keele University and Durham University Joint Report.
23. Kwon, W., Lee, M., & Back, K.-J. (2020). Exploring the underlying factors of customer value in restaurants: A machine learning approach. *International Journal of Hospitality Management*, 91, 102643. <https://doi.org/10.1016/j.ijhm.2020.102643>
24. Kwon, W., Lee, M., & Bowen, J. T. (2022). Exploring customers' luxury consumption in restaurants: A combined method of topic modeling and three-factor theory. *Cornell Hospitality Quarterly*, 63(1), 66–77. <https://doi.org/10.1177/19389655211037667>
25. Laureate, C. D. P., Buntine, W., & Linger, H. (2023). A systematic review of the use of topic models for short text social media analysis. *Artificial Intelligence Review*, 56, 14223–14255. <https://doi.org/10.1007/s10462-023-10471-x>
26. Law, R., Lin, K. J., Ye, H., & Fong, D. K. C. (2024). Artificial intelligence research in hospitality: A state-of-the-art review and future directions. *International Journal of Contemporary Hospitality Management*, 36(6), 2049–2068. <https://doi.org/10.1108/IJCHM-02-2023-0189>
27. Li, W., Guo, K., Shi, Y., Zhu, L., & Zheng, Y. (2018). DWWP: Domain-specific new words detection and word propagation system for sentiment analysis in the tourism domain. *Knowledge-Based Systems*, 146, 203–214. <https://doi.org/10.1016/j.knosys.2018.02.004>
28. Liu, H., Jayawardhena, C., Shukla, P., Osburg, V.-S., & Yoganathan, V. (2024). Electronic word of mouth 2.0 (eWOM 2.0) – The evolution of eWOM research in the new age. *Journal of Business Research*, 176, 114587. <https://doi.org/10.1016/j.jbusres.2024.114587>
29. Luo, J. M., Vu, H. Q., Li, G., & Law, R. (2020). Topic modelling for theme park online reviews: Analysis of Disneyland. *Journal of Travel & Tourism Marketing*, 37(2), 272–285. <https://doi.org/10.1080/10548408.2020.1740138>
30. Luo, Y., He, J., Mou, Y., Wang, J., & Liu, T. (2021). Exploring China's 5A global geoparks through online tourism reviews: A mining model based on machine learning approach. *Tourism Management Perspectives*, 37, 100769. <https://doi.org/10.1016/j.tmp.2020.100769>
31. Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., ... & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2-3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
32. Marcolin, C. B., Becker, J. L., Wild, F., Behr, A., & Schiavi, G. (2021). Listening to the voice of the guest: A framework to improve decision-making processes with text data.

- International Journal of Hospitality Management*, 94, 102853. <https://doi.org/10.1016/j.ijhm.2020.102853>
33. Mazarura, J., & De Waal, A. (2016). A comparison of the performance of latent Dirichlet allocation and the Dirichlet multinomial mixture model on short text. *2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)* (pp. 1–6). Stellenbosch, South Africa: IEEE. <https://doi.org/10.1109/RoboMech.2016.7813155>
34. Mirzaalian, F., & Halpenny, E. (2021). Exploring destination loyalty: Application of social media analytics in a nature-based tourism setting. *Journal of Destination Marketing & Management*, 20, 100598. <https://doi.org/10.1016/j.jdmm.2021.100598>
35. Nguyen, V.-H., & Ho, T. (2023). Analysing online customer experience in hotel sector using dynamic topic modelling and net promoter score. *Journal of Hospitality and Tourism Technology*, 14(2), 258–277. <https://doi.org/10.1108/JHTT-04-2021-0116>
36. Núñez, J. C. S., Gómez-Pulido, J. A., & Ramírez, R. R. (2024). Machine learning applied to tourism: A systematic review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 14(5), e1549. <https://doi.org/10.1002/widm.1549>
37. Papilloud, C., & Hinneburg, A. (2018). *Qualitative Textanalyse mit Topic-Modellen: Eine Einführung für Sozialwissenschaftler* [Qualitative text analysis with topic models: An introduction for social scientists]. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-21980-2>
38. Park, E., Chae, B., Kwon, J., & Kim, W.-H. (2019). The effects of green restaurant attributes on customer satisfaction using the structural topic model on online customer reviews. *Sustainability*, 12(7), 2843. <https://doi.org/10.3390/su12072843>
39. Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58(4), 1064-1082. <https://doi.org/10.1111/ajps.12103>
40. Sánchez-Franco, M. J., & Aramendia-Muneta, M. E. (2023). Why do guests stay at Airbnb versus hotels? An empirical analysis of necessary and sufficient conditions. *Journal of Innovation & Knowledge*, 8(3), 100380. <https://doi.org/10.1016/j.jik.2023.100380>
41. Sanchez-Franco, M. J., Cepeda-Carrion, G., & Roldán, J. L. (2019). Understanding relationship quality in hospitality services: A study based on text analytics and partial least squares. *Internet Research*, 29(3), 478–503. <https://doi.org/10.1108/IntR-12-2017-0531>
42. Shafqat, W., & Byun, Y.-C. (2020). A recommendation mechanism for under-emphasized tourist spots using topic modeling and sentiment analysis. *Sustainability*, 12(1), 320. <https://doi.org/10.3390/su12010320>
43. Shang, Z., & Luo, J. (2022). Topic modeling for hiking trail online reviews: Analysis of the Mutianyu Great Wall. *Sustainability*, 14(6), 3246. <https://doi.org/10.3390/su14063246>
44. Shang, Z., Luo, J. M., & Kong, A. (2022). Topic modelling for ski resorts: An analysis of experience attributes and seasonality. *Sustainability*, 14(6), 3533. <https://doi.org/10.3390/su14063533>
45. Sim, Y., Lee, S. K., & Sutherland, I. (2021). The impact of latent topic valence of online reviews on purchase intention for the accommodation industry. *Tourism Management Perspectives*, 40, 100903. <https://doi.org/10.1016/j.tmp.2021.100903>
46. Srinivas, S., & Ramachandiran, S. (2024). Passenger intelligence as a competitive opportunity: Unsupervised text analytics for discovering airline-specific insights from online reviews. *Annals of Operations Research*, 333, 1045–1075. <https://doi.org/10.1007/s10479-022-05162-9>

47. Sutherland, I., & Kiatkawsin, K. (2020). Determinants of guest experience in Airbnb: A topic modeling approach using LDA. *Sustainability*, 12(8), 3402. <https://doi.org/10.3390/su12083402>
48. Taecharungroj, V. (2023). Experiential brand positioning: Developing positioning strategies for beach destinations using online reviews. *Journal of Vacation Marketing*, 29(3), 313–330. <https://doi.org/10.1177/13567667221095588>
49. Tang, F., Yang, J., Wang, Y., & Ge, Q. (2022). Analysis of the image of global glacier tourism destinations from the perspective of tourists. *Land*, 11(10), 1853. <https://doi.org/10.3390/land11101853>
50. Tang, J., Meng, Z., Nguyen, X., Mei, Q., & Zhang, M. (2014). Understanding the limiting factors of topic modeling via posterior contraction analysis. *Proceedings of the 31st International Conference on Machine Learning* (pp. 190-198). Beijing, China: JMLR: W&CP.
51. Twil, A., Bidan, M., Bencharef, O., Kaloun, S., & Safaa, L. (2021). Exploring destination's negative e-reputation using aspect based sentiment analysis approach: Case of Marrakech destination on TripAdvisor. *Tourism Management Perspectives*, 40, 100892. <https://doi.org/10.1016/j.tmp.2021.100892>
52. Vargas-Calderón, V., Moros Ochoa, A., Castro Nieto, G. Y., & Camargo, J. E. (2021). Machine learning for assessing quality of service in the hospitality sector based on customer reviews. *Information Technology & Tourism*, 23(3), 351–379. <https://doi.org/10.1007/s40558-021-00207-4>
53. Viñán-Ludeña, M. S., & de Campos, L. M. (2022). Analyzing tourist data on Twitter: A case study in the province of Granada at Spain. *Journal of Hospitality and Tourism Insights*, 5(2), 435–464. <https://doi.org/10.1108/JHTI-11-2020-0209>
54. Wang, J., Li, Y., Wu, B., & Wang, Y. (2021). Tourism destination image based on tourism user generated content on internet. *Tourism Review*, 76(1), 125–137. <https://doi.org/10.1108/TR-04-2019-0132>
55. Wen, H., Park, E., Tao, C.-W., Chae, B., Li, X., & Kwon, J. (2020). Exploring user-generated content related to dining experiences of consumers with food allergies. *International Journal of Hospitality Management*, 85, 102357. <https://doi.org/10.1016/j.ijhm.2019.102357>
56. Wu, L., Yang, W., Gao, Y. (L.), & Ma, S. (D.). (2022). Feeling luxe: A topic modeling × emotion detection analysis of luxury hotel experiences. *Journal of Hospitality & Tourism Research*, 47(8), 1425–1452. <https://doi.org/10.1177/10963480221103222> (Original work published 2023).
57. Xu, J., Hsiao, A., Reid, S., & Ma, E. (2023). Working with service robots? A systematic literature review of hospitality employees' perspectives. *International Journal of Hospitality Management*, 113, 103523. <https://doi.org/10.1016/j.ijhm.2023.103523>
58. Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. *International World Wide Web Conference* (pp. 1445–1456). Rio de Janeiro, Brazil: ACM. <https://doi.org/10.1145/2488388.2488514>
59. Zhang, J. (2019). What's yours is mine: Exploring customer voice on Airbnb using text-mining approaches. *Journal of Consumer Marketing*, 36(5), 655–665. <https://doi.org/10.1108/JCM-02-2018-2581>
60. Zolfaghari, A., & Choi, H. C. (2023). Elevating the park experience: Exploring asymmetric relationships in visitor satisfaction at Canadian national parks. *Journal of Outdoor Recreation and Tourism*, 43, 100666. <https://doi.org/10.1016/j.jort.2023.100666>
61. Zou, L., & Song, W. W. (2016). LDA-TM: A two-step approach to Twitter topic data clustering. *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 342–347). Chengdu: IEEE. <https://doi.org/10.1109/ICCCBDA.2016.7529581>