International Journal of Advances in Applied Sciences (IJAAS)
Vol. 13, No. 1, March 2024, pp. 63~72
ISSN: 2252-8814, DOI: 10.11591/ijaas.v13.i1.pp63-72

Artificial intelligence assisted insights into Bali's destination image: sentiment and thematic analyses of TripAdvisor reviews

Putu Chris Susanto[1], Putu Wida Gunawan[2], I Gde Dhika Widarnandana[3]
[1]Department of Management, Faculty of Business, Tourism, Education, and Humanities, Universitas Dhyana Pura, Badung, Indonesia
[2]Department of Computer Science, Faculty of Technology and Informatics, Universitas Dhyana Pura, Badung, Indonesia
[3]Department of Psychology, Faculty of Health and Science, Universitas Dhyana Pura, Badung, Indonesia

Article Info

Article history:

Received Nov 11, 2023
Revised Nov 27, 2023
Accepted Dec 9, 2023

Keywords:

Destination image
Natural language processing
Sentiment analysis
Thematic content analysis
Tourism marketing

ABSTRACT

This study applies sentiment and thematic content analyses based on natural language processing (NLP) to gain valuable insights into the perceived image of Bali as a tourist destination. This study addresses the gap in how to realize the benefits of big data analytics in applied research, by using more approachable tools for researchers with limited programming skills and coding experience. A total of 6,800 TripAdvisor reviews of Bali's top 12 tourist attractions between May 2019 and April 2023 were scrapped. The authors used Bardeen.ai for data mining and Atlas.ti for qualitative data analyses. Sentiment analysis revealed an overwhelmingly positive sentiment (70.4%) towards Bali's tourist attractions, indicating a positive destination image. Post-pandemic tourists tend to express more positive sentiments in their reviews compared to pre-pandemic. Thematic content analysis indicated that positive sentiments are strongly related to satisfaction, positive experiences, enjoyment, and excitement, while environmental concerns and dissatisfaction are potentially harmful to Bali's destination image. The study provides valuable insights into tourists' emotional sentiments, perceptions, and thematic patterns of behavior, which can inform tourism marketers and destination strategists, and contribute to the larger discussion of utilizing big data analytics in tourism marketing research.

Corresponding Author:

Putu Chris Susanto
Department of Management, Faculty of Business, Tourism, Education, and Humanities
Universitas Dhyana Pura
Raya Padang Luwih Street, Tegaljaya, Dalung, Kuta Utara, Badung, Bali, 80361, Indonesia
Email: chris.susanto@undhirabali.ac.id

1. INTRODUCTION

The concept of destination image plays a pivotal role in the field of destination marketing, influencing and even shaping tourists' perceptions and experiences. Tourists' perceptions of a destination's macro and micro images serve as significant precursors to their evaluations of the quality of the destination, affecting the destination's perceived value [1]. A positive destination image has been found to significantly enhance trust and emotional attachment to a destination, potentially contributing to a destination's marketing appeal and competitiveness in the global tourism market [2].

Tourism, with its intangible and experiential nature, often involves high levels of risk perception among its consumers (i.e., tourists). A strong and positive destination image could play a role in mitigating tourists' perceived risks, while simultaneously enhancing behavioral intentions (e.g., intention to revisit and willingness to recommend the destination) [3]. In this era of real-time information, tourism consumers increasingly rely on electronic word-of-mouth (eWOM) to make more informed decisions through learning from other tourists' past experiences. Online travel review (OTR) platforms, such as TripAdvisor, have become increasingly important for sharing information, shared experiences, and reviews. In the case of TripAdvisor, one study found a positive relationship between eWOM and the users' satisfaction and trust in this OTR platform [4], while another study pointed out that the information quality, website quality, and customer satisfaction associated with TripAdvisor influence users' trust in the platform, and that trust is a predictor of eWOM [5].

The importance of OTR in shaping tourist decisions is well-documented. TripAdvisor's user-generated content (UGC) provides a rich data source for analyzing tourists' experiences, preferences, and even behavior [6]. The use of big data, web analytics, and machine learning in destination marketing is emerging. Analyses of tourism eWOM and UGC in the form of OTRs, such as sentiment analysis and thematic content analysis [7], offer novel methodologies in capturing authentic experiences and unfiltered insights from the tourists' perspectives, complementing traditional survey methods that may not fully capture the tourists' sentiments and lived experiences. TripAdvisor is singled out because it is the world's leading global travel platform and the leading online travel community site—with over 460 million monthly active users and over 860 million reviews on 8.7 million attractions/experiences, accommodations, and restaurants worldwide [8].

Previous studies have shown the potential of using reviews on TripAdvisor to understand tourists' sentiments, preferences, and perceptions of various destinations [6], [8]. Despite the many potential benefits of using big data analytics in tourism marketing research, this type of research in the Indonesian context is still lacking. Studies published are largely still limited to sentiment analysis in the contexts of hotel guest experience and satisfaction [9], the sentiments of restaurant patrons [10], and individual tourist attractions [11]. One study used sentiment analysis and thematic content analysis to explore the most prominent emotions uttered by domestic tourists at four select attractions across Indonesia [12]. There is still a gap in the use of OTRs for analyzing generalized sentiments and reviewing emerging themes related to tourist marketing in Indonesia.

This phenomenon is also true in the case of Bali, Indonesia's premier tourist destination. While Bali has been extensively studied as a tourist destination [13], there is a lack of comprehensive analysis of tourist reviews on TripAdvisor that delve into the sentiments, preferences, and experiences expressed by tourists. This gap hinders the understanding of the factors influencing tourist perceptions regarding Bali as a destination, and how these perceptions in turn affect Bali's destination image. Even though Bali is ranked as the second most popular destination in the world in 2023 by TripAdvisor, few studies have realized the potential of using big data (in the form of TripAdvisor reviews) for destination marketing research in the Balinese context [14], [15].

The lack of studies taking advantage of big data in the form of eWOM and UGC for tourism marketing research in the context of Indonesia, and more specifically Bali, indicates that this field is still emerging. Research-based on big data analytics offers a novel way of acquiring data and information in marketing [16]. Yet, its application is still limited due to the programming skills and know-how in data analytics often required—which many marketing and tourism researchers lack. As such, there remain gaps in understanding consumer-generated data in the digital age by using approachable methodologies for analyzing online reviews.

While the language models and sentiment analyses (including domain-specific models like TourBERT) can be quite powerful [17], they can be highly technical to be applied by less tech-savvy marketing and tourism researchers [18]. Many sentiment analysis tools require substantial programming or coding skills to implement and customize. This technical barrier can limit the accessibility of sentiment analysis for applied researchers with limited programming knowledge or skills. Developing and training sentiment analysis models can be computationally intensive, requiring access to powerful hardware and software resources, which is why many applied researchers still rely on semi-manual techniques in analyzing online consumer data.

To bridge the gap, this study proposes a novel approach that employs sentiment analysis and thematic content analysis to applied marketing research using automated tools and data processing software that are less intimidating for researchers with limited programming skills. In this study, the authors explore tourists' perception of Bali as a destination based on reviews posted on TripAdvisor, focusing on 12 of the most popular tourist attractions in Bali. This study aims to apply data mining and machine learning through sentiment and thematic content analyses based on natural language processing (NLP) to gain an understanding of tourist perceptions of Bali's destination image. NLP is powerful because focuses on the interactions between computers and human languages, combining artificial intelligence (AI), cognitive science, and linguistics [19]. TripAdvisor traveler reviews as a form of eWOM were chosen as the object of research, as the reviews on the OTR platform are done intentionally (by tourists with positive and negative travel experiences), and uploaded on platforms that specifically discuss travel experiences (as opposed to generalized platforms such as Facebook and Instagram). Since TripAdvisor does not provide any means to review a destination in general, reviews of the 12 most popular tourist attractions are used as a proxy for Bali's overall destination image.

In this study, the authors seek to actualize the benefits of NLP for applied sciences. The study employed easy-to-use automation tools like Bardeen.ai for data collection and processing and Atlas.ti qualitative data analysis (QDA) tool for analysis—both of which require minimum technical coding skills and computational power. This approach helps make big data analysis more accessible to a broader range of researchers. It also allows the data collection, pre-processing, and analysis to be more simplified. The novel aspect of this study is the use of more accessible big data analytics methods for researchers with limited programming skills, to simplify the technical complexities associated with big data analytics and allow more focus on gaining insights from the data. The study's finding is expected to contribute to the existing literature by providing insights into the emotional responses, sentiments, and thematic patterns by reviewing tourists' lived experiences as they share their travel experiences and perceived image of Bali as a tourist destination. The findings should have practical implications for policymakers and destination marketers while contributing to the broader field of tourism marketing research in the digital era.

2. RESEARCH METHOD

The methods employed in this study aim to address the aforementioned research objectives and knowledge gaps. This study used computer-assisted, NLP-based machine learning to analyze and understand human language automatically, thus allowing the authors to extract the meaning contained in large amounts of textual data [20]. The study was conducted in five distinct stages.

In the first stage, the authors collected a substantial number of review data from TripAdvisor, the world's largest travel platform, which were written in English. Bardeen, a generative AI tool for automation, was used to scrape the data. Bardeen.ai is a free online task management tool that allows users to automate repetitive processes, including scraping data from various online platforms (e.g., TripAdvisor) [21]. The authors used Bardeen's "Scraper" tool in its builder interface to scrape the data on active TripAdvisor review tabs, then automated the process by adding each new review as a new row in Google Sheets.

Review data from 12 select popular tourist attractions in Bali were scrapped. As aforementioned, TripAdvisor does not provide reviews on destinations per se. As such, 12 popular tourist attractions were used as the proxy to analyze the overall destination image, covering both coastal and inland sites, natural and man-made ranging from beaches, temples, rice terraces, mountains, nature sanctuaries, and theme parks). The attractions included eight coastal sites (i.e., Bali Safari and Marine Park, Kelingking Beach, Kuta Beach, Nusa Dua Beach, Seminyak Beach, Tanah Lot Temple, Uluwatu Temple, and Waterbom Bali), and four inland sites (i.e., Jatiluwih Rice Terraces, Mount Batur, Tegallalang Rice Terraces, and Ubud Monkey Forest). The scraping process followed TripAdvisor's terms of service and ethical guidelines for data collection.

The authors collected 6,815 popular tourist attractions review data from 12 Bali's most popular tourist attractions on TripAdvisor. These reviews were posted between 1 May 2019 and 30 April 2023. The time constraint was specifically chosen to represent two distinct periods in Bali's tourism: the pre-pandemic period before 30 March 2020 (the official closing of Bali's International borders for visitors), and the post-pandemic period after 1 April 2020. Upon data cleaning, 6,800 review data were deemed to be valid for further analysis. The data collected included the review text, date of review, user rating, and user location

In the second stage, the collected data underwent a pre-processing procedure to ensure their relevance to the research. This involved cleaning the data to remove duplicate reviews, missing values, and irrelevant information. Subsequently, the data was converted into a text format suitable for analysis. The authors performed data preprocessing that involved cleaning the data, tokenizing (paragraphs), removing punctuation marks, and converting emojis using RapidMiner. Since the authors used Atlas.ti as a robust QDA tool, it does not require stemming and lemmatization in the process of data pre-processing. Atlas.ti also accepts data using regular text formats, which the authors uploaded to the software. The normalized data, following the stop-word removal process, was then used to ensure consistency in analysis. Each review was assigned a unique identifier and categorized based on the tourist attraction to which it pertained.

Subsequently, in the third stage, the authors used Atlas.ti to conduct sentiment and thematic content analyses. Atlas.ti is a robust QDA tool that offers various text analysis packages to automate the coding and content analysis process based on AI [22]. In conducting the sentiment analysis, the authors used Atlas.ti's pre-trained advanced English language model to categorize the reviews into three categories: negative, neutral, and positive. This classification was based on the presence of specific keywords and phrases that are indicative of the reviewer's sentiment, which was done by Atlas.ti's AI-assisted automated sentiment analysis. After selecting the documents to be analyzed and running the sentiment analysis function, the authors only had to verify and apply the proposed sentiment codes to the respective documents. Then, using the code-document analysis tool a table and a Sankey diagram were created. This can be used to visualize the connections between the attractions and tourist sentiments, for mapping input-output flow or linkage between two related or juxtaposed concepts [23].

In the fourth stage, the authors used SPSS to conduct cross-tabulation analysis using Chi-square to determine whether there are statistically significant associations in the cross-tabulation. Two different crosstabs were conducted. First, cross-tabulation between the type of destinations (i.e., coastal and inland) and sentiment (i.e. negative, neutral, and positive) was created. Second, cross-tabulation between periods (i.e., pre-pandemic and post-pandemic) and sentiment (i.e. negative, neutral, and positive) was also created.

In the final stage, the authors conducted thematic content analysis using Atlas.ti's AI coding capabilities. Upon selecting the documents to analyze, Atlas.ti carried out an automated AI-assisted coding process. Atlas.ti was able to facilitate the identification of themes and patterns in the qualitative data (i.e., TripAdvisor textual reviews). The software analyzed the textual data and generated codes representing the main themes and sub-themes emerging from the reviews [22]. While AI coding was a fully automated process, the authors still had to double-check that all the codes and themes were correctly identified and categorized. The software provided suggestions on how the codes could be grouped into categories based on their relevance. Subsequently, the frequency of each code was analyzed to identify the emerging themes. While the entire sentiment analysis process was automated, thematic content analysis cannot be fully considered a topical modeling process, as is not yet considered a fully automated and algorithmic method.

From the thematic content analysis, the authors created two tables. The first table is a cross-tabulation of emerging themes and sentiments, and the second is a cross-tabulation of emerging themes and attractions. The findings were subsequently interpreted in the context of the research objectives and the existing literature on tourism marketing research. The analysis used a combination of descriptive statistics (e.g., frequencies and percentages), and inferential statistics (e.g., chi-square tests), to assess the significance of the findings.

3. RESULTS AND DISCUSSION

In this study, the data source was travelers' eWOM in the form of TripAdvisor reviews. Tourist reviews were chosen as the object of research because the reviews on the platform were done intentionally, and uploaded on a platform that specifically discusses the travel experience—i.e., not on social media platforms where the content is mixed. The authors only used reviews written in the English language. Popular tourist attractions were chosen as the unit of analysis as a proxy for the overall destination image of Bali. A period of three years (covering pre- and post-pandemic periods) was also chosen to ensure the adequacy of data.

3.1. Sentiment analysis

Of the 12 popular tourist attractions in Bali, Ubud Monkey Forest had the highest number of reviews within the period (2,052 reviews or 30.2%), followed by Tegallalang Rice Terraces (1,142 reviews or 16.8%). The attraction with the least number of reviews was Jatiluwih Rice Terraces (108 reviews or 1.6%). Sentiment analysis on all 12 popular attractions indicated an overall positive sentiment towards tourist attractions in Bali (70.4%), while only 13.4% of the sentiments were negative and 16.3% were neutral (Table 1). The overall positive sentiment is encouraging for Bali's tourism industry. Visitors generally have positive experiences, translating into a more positive brand image of Bali as a tourism destination which bodes well for attracting even more tourists. Studies have shown that visitor experience is closely linked to destination image and behavioral intention [24], and that positive destination experience contributes to tourist satisfaction, revisit intention, and willingness to recommend [25].

Of the 6,800 reviews, 47.7% were written for coastal attractions and 52.3% for inland attractions. This suggests that coastal and inland attractions garnered a similar proportion of reviews. The findings also indicated that 79.0% were posted pre-COVID-19 pandemic and 21.0% were posted post-pandemic, suggesting that understanding sentiment changes during and after the pandemic may help in adapting tourism strategies. This supports the need to extract insights and understanding of tourists' sentiments and emotions through shared postings and reviews during the pandemic time by using big data analysis [26], to better inform tourism marketing efforts and messages post-pandemic [27].

Seven attractions had higher-than-average positive sentiments, with Waterbom Bali having the highest proportion of positive sentiments (85.9%). Five attractions had lower-than-average positive sentiments, with Kuta Beach having the lowest proportion of positive sentiments (57.3%). Waterbom Bali also had the lowest proportion of negative sentiments a mere 7.0%, while Kuta Beach had the highest proportion of negative sentiments a whopping 21.7%. Waterbom Bali stands out with the highest positive sentiment while having the lowest proportion of negative sentiment. Conversely, Kuta Beach had the lowest proportion of positive sentiments and the highest proportion of negative sentiments (Figure 1). Though both attractions are in close proximity to one another, both garnered markedly different responses from visitors. Waterbom Bali is a privately operated man-made water-themed park, while Kuta Beach is a publicly operated natural attraction. This warrants further study into tourist experiences in each respective site. This may reflect the dichotomy between publicly vs. privately operated spaces, as one study noted that while public sites are preferred for their relatively easy access, privately operated sites are preferred for better infrastructure and safety [28].

From Table 1 and Figure 1, it is evident that the two attractions that received the highest number of reviews were Ubud Monkey Forest (2,052) and Tegallalang Rice Terraces (1,142). Together, they account for 47% of the reviews for the determined period. This is perhaps due to their geographic proximity in the Ubud area—both of which visitors can experience in a full-day or even half-day Ubud tour. They are also popular among tour guides and tour operators offering packaged inland tours [29]. The sites were made even more popular by the book and subsequent film "Eat Pray Love" [30].

Interestingly, while Jatiluwih Rice Terraces is a registered UNESCO World Heritage site [31], it is less popular than Tegallalang Rice Terraces. Again, this is perhaps due to the lack of proximity between Jatiluwih and other popular tourist sites, making it less likely for visitors to arrange a combined trip with other interesting sites (as opposed to the proximity between Ubud, Tegallalang, and other sites such as Tampaksiring temple and Kintamani/Mount Batur). One study used big data collected from mobile phone locations of tourists engaging in day trips, suggesting that tourists tend to prefer day-trip chains (i.e., sites/locations that are close to one another, so that they can visit multiple sites in one day) [32]. This is also seemingly true in the case of Bali, although further study is recommended.

Further, the authors conducted a cross-tabulation analysis to determine whether there are significant differences in the sentiments between coastal and inland attractions, as well as between reviews posted pre-pandemic and post-pandemic. Table 2 shows the cross-tabulation results juxtaposing attraction types and sentiments. The Chi-square test yielded a calculated value of 12.789, representing the difference between the expected (theoretical) values and the observed values [33], with a critical p-value of 0.002 ($p < 0.05$). This signifies a significant relationship between attraction type and sentiment. Coastal attractions are more likely to garner positive sentiment and less likely to garner neutral sentiments, compared to inland attractions. Coastal attractions appear to elicit more positive sentiments and fewer neutral sentiments. One potential explanation for this is congestion in certain attractions, as noted in another study in Spain [34]. This insight can be valuable in understanding how different types of attractions influence visitor sentiments, which may have implications for destination management and marketing strategies for Bali as a tourism destination.

A cross-tabulation analysis juxtaposing review periods and the sentiments indicates that tourists are more likely to write positive reviews and less likely to write neutral revires post-COVID-19 pandemic (Table 3). The Chi-Square test yielded a calculated value of 19.293, with a critical p-value of 0.000 ($p<0.05$) signifying a significant relationship between the review period and sentiment. This finding suggests a notable shift in tourist sentiments between pre-pandemic and post-pandemic periods. Post-pandemic, tourists are more inclined to express positive sentiments in their reviews, while neutral sentiments have decreased. This finding supports one study's finding that restaurant patrons tend to share more positive sentiments post-pandemic compared to pre-pandemic [35]. This shift may reflect changing tourist experiences and perceptions in response to the pandemic's impact, which warrants further investigation into tourists' post-pandemic sentiments and behavior.

In brief, the sentiment and cross-tabulation analyses provide valuable insights for understanding sentiment patterns depicting tourists' overall emotional responses to popular tourist attractions in Bali. The insights can be used to indicate the overall sentiment on the image of Bali as a tourist destination, enhance the island's tourism marketing strategies, improve visitor experiences, and tailor efforts to specific attractions. This is expected to benefit Bali's tourism industry, just like it has for other destinations such as Hong Kong (China) [36], Granada (Spain) [37], Cilento (Italy) [38], and Marrakech (Morocco) [39].

Table 1. Summary of sentiment analysis for top 12 Bali attractions

No.	Attraction	Type	Reviews	Negative (%)	Neutral (%)	Positive (%)
1.	Ubud Monkey Forest	Inland	2,052	11.8	16.7	71.5
2.	Tegallalang Rice Terraces	Inland	1,142	14.8	18.3	66.9
3.	Waterbom Bali	Coastal	630	7.0	7.1	85.9
4.	Kuta Beach	Coastal	585	21.7	21.0	57.3
5.	Uluwatu Temple	Coastal	483	12.6	15.3	72.0
6.	Tanah Lot Temple	Coastal	429	10.0	13.5	76.5
7.	Nusa Dua Beach	Coastal	353	13.9	14.7	71.4
8.	Kelingking Beach	Coastal	334	14.7	19.8	65.6
9.	Bali Safari and Marine Park	Coastal	259	15.1	11.6	73.4
10.	Mount Batur	Inland	253	17.0	24.9	58.1
11.	Seminyak Beach	Coastal	172	18.6	14.5	66.9
12.	Jatiluwih Rice Terraces	Inland	108	10.2	15.7	74.1
	Total and averages		6,800	13.4	16.3	70.4

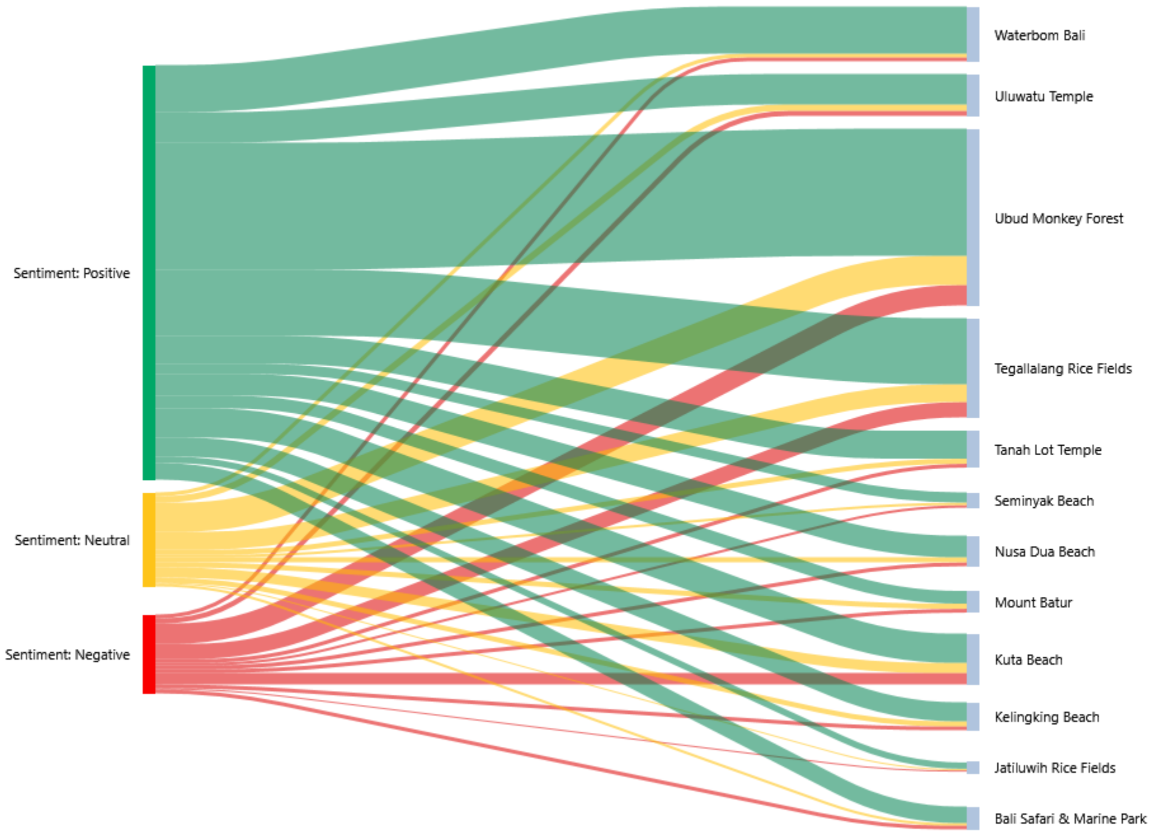

Figure 1. Sankey diagram on sentiment analysis for top 12 Bali attractions

Table 2. Cross-tabulation of attraction type and sentiment

		Sentiment		
		Negative (%)	Neutral (%)	Positive (%)
Type	Coastal	13.7	14.6	71.7
	Inland	13.1	17.8	69.1
	Total	13.4	16.3	70.4

Table 3. Cross-tabulation of review period and sentiment

		Sentiment		
		Negative (%)	Neutral (%)	Positive (%)
Period	Pre-pandemic	13.6	17.2	69.2
	Post-pandemic	12.4	12.8	74.8
	Total	13.4	16.3	70.4

3.2. Thematic content analysis

The AI coding using Atlas.ti initially yielded 3,821 unique codes. Upon evaluation and processing involving code merging and reclassification, the authors were able to narrow down these codes into 3,480 remaining unique codes. Of these codes, 30 appeared in 100 or more quotes from the 6,800 reviews analyzed in this study. The 30 codes consisted of individual codes (i.e., stand-alone) and categorical codes (i.e., containing sub-categories of code underneath). After processing and renaming the codes as necessary for content analysis, the authors then conducted a co-occurrence analysis juxtaposing the 30 most frequently appearing themes (from the AI coding and subsequent processing) with tourists' sentiments. The results of the co-occurrence analysis are shown in Table 4, sorted by 'groundedness' (i.e., the number of quotations coded by any particular code), which is used to determine the significance of certain themes [40].

As shown in Table 4, the theme 'travel and tourism' was found to be the one with the most groundedness, which means it appeared in the most quotations across 6,800 tourist reviews. The theme that appeared the second-most in review quotations was 'nature', signifying that Bali is still largely a travel destination that is known for its nature (including beaches, lakes, mountains, rice terraces, etc.). Compared to comparable island destinations such as Singapore and Hong Kong, Bali is still known for its natural attractions—rather than man-made ones [41]. After all, 'beauty' also emerged in the top eight themes.

In terms of environmental sustainability, 236 quotations were tagged 'environmental concerns', with sub-themes ranging from environmental awareness, degradation, plastic pollution, and water quality, to noise pollution (Table 4). As tourists voicing 'environmental concerns' are likely to perceive Bali's tourist attractions as neutral (24%) and even negative (30%), serious efforts must be taken to ensure sustainability. This is in line with a study suggesting the importance of environmental concerns in ensuring sustainable growth while considering economic and social concerns [42].

As shown in Table 4, positive sentiment is (unsurprisingly) correlated to 'satisfaction' (90%), 'positive experience' (89%), 'recommendation' (87%), 'enjoyment' (86%), 'appreciation/admiration' (85%), and 'excitement' (83%). Positive sentiment is also found to be highly correlated to 'hospitality' (92%), 'cleanliness' (88%), and 'family-friendly' (84%). These findings suggest that when tourists express positive sentiments, it is often because they are satisfied with their overall experience at the destination and that they have had positive experiences when visiting the attraction/destination. This is in line with the assertion that an essential part of tourists' positive experience is satisfaction, both of which are related to positive sentiment [43].

Table 4. Co-occurrence analysis of emerging themes and sentiments

No.	Theme	Sentiment			Total Q
		Negative (%)	Neutral (%)	Positive (%)	
1.	Travel and tourism	10	19	71	2148
2.	Nature	6	15	79	1625
3.	Positive experience	3	8	89	1537
4.	Tourist experience	16	19	65	1461
5.	Enjoyment	4	10	86	1354
6.	Tourist attractions	11	17	72	1046
7.	Recreation and leisure	6	15	79	954
8.	Beauty	5	14	81	828
9.	Caution	17	23	60	799
10.	Recommendation	5	9	87	773
11.	Appreciation/admiration	4	11	85	725
12.	Safety	20	22	58	716
13.	Cultural experience	10	19	70	709
14.	Adventure	8	21	71	610
15.	Animals and wildlife	10	15	74	606
16.	Costs and pricing	18	20	63	545
17.	Excitement	6	10	83	524
18.	Crowdedness	23	19	58	400
19.	Satisfaction	3	7	90	346
20.	Photography	9	16	75	344
21.	Family-friendly	6	10	84	340
22.	Food and beverages	4	18	79	317
23.	Dissatisfaction	41	20	39	307
24.	Beach	14	20	66	305
25.	Accessibility	16	28	57	257
26.	Hospitality	4	5	92	254
27.	Environmental concerns	30	24	46	236
28.	Cleanliness	6	6	88	234
29.	Desire for improvement	14	27	59	110
30.	Annoyance	32	34	34	100

3.3. General discussion

Overall, the findings highlighted the significance of Bali's natural beauty, and themes like cleanliness and hospitality in shaping tourist experiences and sentiments (and thus the island's destination image), while also noting themes such as environmental concerns and dissatisfaction as potentially harmful to Bali's image. To sustain the image as a destination with abundant natural beauty, a place for enjoyment, as well as a destination for cultural experiences, Bali as a tourism destination must make active strides towards protecting the island's natural beauty and ensuring its environmental sustainability. This must be done as the island faces serious environmental crises arising from overdevelopment and over-tourism including water shortage, converted use of agricultural land, displacement, and pollution [47]. These issues can have an adverse impact on Bali's sustainability as well as the island's destination image.

Further, the findings from the thematic content analysis support previous research asserting that big data analysis in tourism marketing can help improve decision-making for destination managers, and formulate marketing strategies with a higher degree of personalization, transparency, and engagement with various stakeholders [48]. Another study suggests that big data can be combined with small data (e.g., visual eWOM, 'ethnography', and qualitative geographic information systems (GIS)) to gain even deeper insights into tourism marketing [49]. One limitation of this study, however, is the potential of 'noise' in the dataset derived from the TripAdvisor review. Although the authors have conducted AI-assisted data cleaning, some noise in the data (e.g., data not cleaned properly, grammatical errors, and typos) could still slightly affect the results of the analyses.

Another limitation is that since this data was collected from OTR (i.e., TripAdvisor), the data may not be representative of the broader population of international tourists. This is because reviews in OTRs tend to involve self-selection, and the review tends to skew towards active TripAdvisor users only. Since the users posting the reviews are self-selected, there is a potential for bias (either positive or polarizing bias) as the users who post reviews tend to be those who have had overwhelmingly positive experiences or disappointingly negative ones. Although OTRs are more specific than social media posts, they can also be overly detailed which could actually skew the analysis towards neutrality (i.e., when a detailed review containing both positive and negative sentiments is categorized as 'neutral' by the software). Further study is needed to address these limitations while filling the gaps identified in this study. Future studies could enhance big data analytics from OTR platforms with qualitative studies to gain a deeper understanding of tourist experiences.

4. CONCLUSION

This study utilized AI-assisted sentiment and thematic content analyses to examine tourist reviews of Bali's top 12 attractions on TripAdvisor, spanning pre and post-COVID-19 pandemic periods. Findings from the sentiment analysis conducted using Atlas.ti software revealed a predominantly positive sentiment (70.4%) towards Bali's tourist attractions, with coastal attractions and post-pandemic reviews being more likely to elicit positive sentiments. This pre- vs. post-pandemic shift highlights the evolving nature of tourist experiences and perceptions in response to the pandemic's impact. The thematic content analysis highlighted the importance of cleanliness, hospitality, and environmental preservation in shaping Bali's positive destination image. The thematic content analysis showed Atlas.ti's AI coding capabilities, revealing that positive sentiments were strongly related to satisfaction, positive experiences, recommendation, enjoyment, and excitement. However, the study is limited by potential 'noise' in the dataset and the non-representative nature of OTRs, which may exhibit bias due to self-selection. Future research should address these limitations and further explore the evolving nature of tourist experiences by combining it with more in-depth qualitative methods. Overall, the insights from this study could be invaluable for destination management and marketing strategies tailored to specific attraction types, as well as to maintain positive brand image.

ACKNOWLEDGEMENTS

The author thanks the Directorate of Research, Technology, and Community Service (DRTPM) at the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for providing the research grant for this study (contract no. 184/E5/PG.02.00.PL/2023; 3583/LL8/AL.04/2023; 002/UNDHIRA-LPPM/PN/E/2023). The authors also thank LPPM Universitas Dhyana Pura for its continued support for emerging researchers.

REFERENCES

[1] B. B. Dedeoğlu, "Shaping tourists' destination quality perception and loyalty through destination country image: The importance of involvement and perceived value," *Tourism Management Perspectives*, vol. 29, pp. 105–117, Jan. 2019, doi: 10.1016/j.tmp.2018.11.006.

[2] A. Al-Ansi and H. Han, "Role of halal-friendly destination performances, value, satisfaction, and trust in generating destination image and loyalty," *Journal of Destination Marketing and Management*, vol. 13, pp. 51–60, Sep. 2019, doi: 10.1016/j.jdmm.2019.05.007.

[3] M. U. Nazir, I. Yasin, and H. H. Tat, "Destination image's mediating role between perceived risks, perceived constraints, and behavioral intention," *Heliyon*, vol. 7, no. 7, p. e07613, Jul. 2021, doi: 10.1016/j.heliyon.2021.e07613.

[4] J.-G. Martínez-Navalón, V. Gelashvili, and A. Gómez-Ortega, "Evaluation of user satisfaction and trust of review platforms: analysis of the impact of privacy and E-WOM in the case of tripAdvisor.," *Frontiers in psychology*, vol. 12, p. 750527, Sep. 2021, doi: 10.3389/fpsyg.2021.750527.

[5] R. Filieri, S. Alguezaui, and F. McLeay, "Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth," *Tourism Management*, vol. 51, pp. 174–185, Dec. 2015, doi: 10.1016/j.tourman.2015.05.007.

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

[7] D. Buzova, S. Sanz-Blas, and A. Cervera-Taulet, "'Tour me onshore': understanding cruise tourists' evaluation of shore

excursions through text mining," *Journal of Tourism and Cultural Change*, vol. 17, no. 3, pp. 356–373, May 2019, doi: 10.1080/14766825.2018.1552277.

[8] L. Xin *et al.*, "Understanding continuous sharing behavior of online travel community users: a case of TripAdvisor," *Journal of Tourism and Cultural Change*, vol. 21, no. 3, pp. 328–343, May 2023, doi: 10.1080/14766825.2023.2170239.

[9] N. D. Wulandari, M. H. Z. Nuri, and L. Kurniasari, "Customers' satisfaction and preferences using sentiment analysis on traveloka: the case of Yogyakarta Special Region hotels," in *Proceedings of the 1st UMGESHIC International Seminar on Health, Social Science and Humanities (UMGESHIC-ISHSSH 2020)*, 2021. doi: 10.2991/assehr.k.211020.058.

[10] Y. A. Singgalen, "Sentiment analysis on customer perception towards products and services of restaurant in Labuan Bajo," *Journal of Information Systems and Informatics*, vol. 4, no. 3, pp. 511–523, Sep. 2022, doi: 10.51519/journalisi.v4i3.276.

[11] E. Widodo, G. D. Nursyafitri, O. N. Putri, and Y. Nurmalasari, "Sentiment analysis of visitors in Borobudur Temple with support vector machine based on TripAdvisor review," in *Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019)*, Paris, France: Atlantis Press, 2020. doi: 10.2991/assehr.k.201010.009.

[12] H. Irawan, G. Akmalia, and R. A. Masrury, "Mining tourist's perception toward Indonesia tourism destination using sentiment analysis and topic modelling," in *Proceedings of the 2019 4th International Conference on Cloud Computing and Internet of Things*, New York, NY, USA: ACM, Sep. 2019, pp. 7–12. doi: 10.1145/3361821.3361829.

[13] I. G. B. R. Utama, P. C. Susanto, C. P. Trimurti, and N. P. D. Krismawintari, "The market potential and financial feasibility of cable car tourist attraction development in Pelaga, Bali, Indonesia," *International Journal of Advances in Applied Sciences*, vol. 11, no. 4, pp. 373–382, Dec. 2022, doi: 10.11591/ijaas.v11.i4.pp373-382.

[14] Nasa Zata Dina, "Tourist sentiment analysis on TripAdvisor using text mining: A case study using hotels in Ubud, Bali," *African Journal of Hospitality, Tourism and Leisure*, vol. 9, no. 2, pp. 1–10, 2020.

[15] D. S. Bimaputra and E. Sutoyo, "Aspect-based sentiment analysis of hotels in Bali on Tripadvisor using BERT Algorithm," *International Journal of Advances in Data and Information Systems*, vol. 4, no. 2, Apr. 2023, doi: 10.25008/ijadis.v4i2.1284.

[16] L. R. Nair, S. D. Shetty, and S. Deepak Shetty, "Streaming big data analysis for real-time sentiment based targeted advertising," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 1, p. 402, Feb. 2017, doi: 10.11591/ijece.v7i1.pp402-407.

[17] V. Arefeva and R. Egger, "When BERT started traveling: TourBERT—A natural language processing model for the travel industry," *Digital*, vol. 2, no. 4, pp. 546–559, Nov. 2022, doi: 10.3390/digital2040030.

[18] K. P. K. and N. S., "Insights to problems, research trend and progress in techniques of sentiment analysis," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 5, pp. 2818–2822, Oct. 2017, doi: 10.11591/ijece.v7i5.pp2818-2822.

[19] A. Lenci and S. Padó, "Editorial: perspectives for natural language processing between AI, linguistics, and cognitive science," *Frontiers in Artificial Intelligence*, vol. 5, p. 1059998, Nov. 2022, doi: 10.3389/frai.2022.1059998.

[20] Y. Kang, Z. Cai, C.-W. Tan, Q. Huang, and H. Liu, "Natural language processing (NLP) in management research: A literature review," *Journal of Management Analytics*, vol. 7, no. 2, pp. 139–172, Apr. 2020, doi: 10.1080/23270012.2020.1756939.

[21] R. Gozalo-Brizuela and E. C. Garrido-Merchán, "A survey of generative AI applications," Jun. 2023, doi: 10.48550/arXiv.2306.02781.

[22] S. Friese, *Qualitative data analysis with ATLAS.ti*, 3rd ed. Thousand Oaks, CA: SAGE Publications Ltd, 2019. doi: 10.4135/9781529799590.

[23] C. Chong *et al.*, "A visualization method of the economic input–output table: Mapping monetary flows in the form of sankey diagrams," *Sustainability*, vol. 13, no. 21, p. 12239, Nov. 2021, doi: 10.3390/su132112239.

[24] S. Piramanayagam, S. Rathore, and P. P. Seal, "Destination image, visitor experience, and behavioural intention at heritage centre," in *Tourism in India*, Routledge, 2021, pp. 35–52. doi: 10.4324/9781003158783-4.

[25] H. Martins, P. Carvalho, and N. Almeida, "Destination brand experience: a study case in touristic context of the peneda-gerês national park," *Sustainability*, vol. 13, no. 21, p. 11569, Oct. 2021, doi: 10.3390/su132111569.

[26] R. K. Mishra, S. Urolagin, J. A. A. Jothi, A. S. Neogi, and N. Nawaz, "Deep learning-based sentiment analysis and topic modeling on tourism during COVID-19 pandemic," *Frontiers in Computer Science*, vol. 3, Nov. 2021, doi: 10.3389/fcomp.2021.775368.

[27] S. Singh, A. Nicely, J. Day, and L. A. Cai, "Marketing messages for post-pandemic destination recovery-A delphi study," *Journal of Destination Marketing and Management*, vol. 23, p. 100676, Mar. 2022, doi: 10.1016/j.jdmm.2021.100676.

[28] A. M. Fayas, M. T. O. V. Peiris, and K. G. P. Kalugalla, "Public vs private sector owned urban public space performance in terms of user perception in city of Colombo," in *14th International Research Conference - FARU 2021*, Faculty of Architecture Research Unit, Dec. 2021, pp. 49–60. doi: 10.31705/FARU.2021.5.

[29] I. G. A. A. D. Susanthi, M. S. D. Muliawan, and A. A. G. Suarjaya, "The language functions used in guiding conversation: pragmatics approach," *RETORIKA: Jurnal Ilmu Bahasa*, vol. 7, no. 1, pp. 1–9, Apr. 2021, doi: 10.22225/jr.7.1.2903.1-9.

[30] E. Park, "The impacts of film tourism on place change and tourist experience: A lesson from Eat Pray Love in Bali, Indonesia," in *Film Tourism in Asia. Perspectives on Asian Tourism*, Singapore: Springer, 2018, pp. 107–123. doi: 10.1007/978-981-10-5909-4_7.

[31] D. P. O. Prasiasa, D. A. D. S. Widari, and I. G. B. R. Utama, "World cultural heritage designated impacts in Jatiluwih Agritourism, Bali, Indonesia," *International Journal of Professional Business Review*, vol. 8, no. 6, p. e02196, Jun. 2023, doi: 10.26668/businessreview/2023.v8i6.2196.

[32] Y. Luo, "Characterizing tourist daily trip chains using mobile phone big data," May 2022, doi: 10.48550/arXiv.2205.14395.

[33] M. A. Fernández-Cásseres and D. Russi-Pulgar, "Categorical variable analyses," in *Translational Sports Medicine*, Elsevier, 2023, pp. 151–156. doi: 10.1016/B978-0-323-91259-4.00043-6.

[34] I. P. Albaladejo and M. González-Martínez, "Congestion affecting the dynamic of tourism demand: evidence from the most popular destinations in Spain," *Current Issues in Tourism*, vol. 22, no. 13, pp. 1638–1652, Aug. 2019, doi: 10.1080/13683500.2017.1420043.

[35] J.-N. Harba, G. Tigu, and A. A. Davidescu, "Exploring consumer emotions in pre-pandemic and pandemic times. a sentiment analysis of perceptions in the fine-dining restaurant industry in Bucharest, Romania," *International Journal of Environmental Research and Public Health*, vol. 18, no. 24, p. 13300, Dec. 2021, doi: 10.3390/ijerph182413300.

[36] Q. Jiang, C.-S. Chan, S. Eichelberger, H. Ma, and B. Pikkemaat, "Sentiment analysis of online destination image of Hong Kong held by mainland Chinese tourists," *Current Issues in Tourism*, vol. 24, no. 17, pp. 2501–2522, Sep. 2021, doi: 10.1080/13683500.2021.1874312.

[37] M. S. Viñán-Ludeña and L. M. de Campos, "Discovering a tourism destination with social media data: BERT-based sentiment analysis," *Journal of Hospitality and Tourism Technology*, vol. 13, no. 5, pp. 907–921, Nov. 2022, doi: 10.1108/JHTT-09-2021-0259.

[38] M. Paolanti *et al.*, "Tourism destination management using sentiment analysis and geo-location information: a deep learning approach," *Information Technology and Tourism*, vol. 23, no. 2, pp. 241–264, Jun. 2021, doi: 10.1007/s40558-021-00196-4.

[39] 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.
[40] A. Gupta, "A qualitative study of management practices and influence on middle managers' morale in public sector banks using ATLAS.ti," in *Computer Supported Qualitative Research. WCQR 2023*, Cham: Springer, 2023, pp. 18–37. doi: 10.1007/978-3-031-31346-2_2.
[41] I. G. B. R. Utama, I. W. R. Junaedi, P. A. N. Putri, and N. P. D. Krismawintari, "Determining the popularity of island tourist destinations using Google trends analysis," *International Journal of Social Science*, vol. 2, no. 2, pp. 1257–1266, Aug. 2022, doi: 10.53625/ijss.v2i2.2131.
[42] E. Purwanda and W. Achmad, "Environmental concerns in the framework of general sustainable development and tourism sustainability," *Journal of Environmental Management and Tourism*, vol. 13, no. 7, p. 1911, Dec. 2022, doi: 10.14505/jemt.v13.7(63).11.
[43] M. Thelwall, "Sentiment analysis for tourism," in *Big Data and Innovation in Tourism, Travel, and Hospitality*, Singapore: Springer Singapore, 2019, pp. 87–104. doi: 10.1007/978-981-13-6339-9_6.
[44] G. Roy and S. Sharma, "Analyzing one-day tour trends during COVID-19 disruption – applying push and pull theory and text mining approach," *Tourism Recreation Research*, vol. 46, no. 2, pp. 288–303, Apr. 2021, doi: 10.1080/02508281.2020.1858614.
[45] V. Ramanujam and P. Kumar, "Role of TripAdvisor in influencing international tourists visiting Puducherry," *International Journal of Technology Marketing*, vol. 16, no. 4, p. 386, 2022, doi: 10.1504/IJTMKT.2022.126267.
[46] I. K. Astawa, M. Suardani, and A. A. P. Suardani, "Kuta beach area, bali: a study on tourism zoning and carrying capacity," *International Journal of Applied Sciences in Tourism and Events*, vol. 3, no. 1, pp. 81–92, 2019, doi: 10.31940/ijaste.v3i1.1236.
[47] C. Warren and A. Wardana, "Sustaining the unsustainable? Environmental impact assessment and overdevelopment in Bali," *Asia Pacific Journal of Environmental Law*, vol. 21, no. 2, pp. 101–125, Nov. 2018, doi: 10.4337/apjel.2018.02.02.
[48] P. Del Vecchio, G. Mele, V. Ndou, and G. Secundo, "Creating value from social big data: Implications for smart tourism destinations," *Information Processing & Management*, vol. 54, no. 5, pp. 847–860, Sep. 2018, doi: 10.1016/j.ipm.2017.10.006.
[49] F. Xu, N. Nash, and L. Whitmarsh, "Big data or small data? A methodological review of sustainable tourism," *Journal of Sustainable Tourism*, vol. 28, no. 2, pp. 144–163, Feb. 2020, doi: 10.1080/09669582.2019.1631318.

BIOGRAPHIES OF AUTHORS

Putu Chris Susanto received his MBA in marketing from the University of Missouri, USA, and his M.Ed. in education from Central Methodist University, USA. Previously, he received his bachelor's degree from the College of the Ozarks USA. He is also a Certified Marketing Analyst and Certified Hospitality Educator. Currently, he is an educator of marketing and entrepreneurship in the Faculty of Business, Tourism, Education, and Humanities at Universitas Dhyana Pura, Bali, Indonesia. His research interests include tourism marketing, consumer behavior, multimodal marketing communication (visual and textual), branding, tourism entrepreneurship, and social entrepreneurship. He is also actively involved in community empowerment projects, including rural tourism development and marketing, eco-entrepreneurship, and community empowerment through entrepreneurial development. He can be contacted at email: chris.susanto@undhirabali.ac.id.

Putu Wida Gunawan holds a master's degree in Computer Science from Universitas Gadjah Mada, Indonesia, having previously completed a bachelor of science at Universitas Udayana, Indonesia. He is a certified Microsoft Specialist. At present, he leads the Department of Computer Science at the Faculty of Technology and Informatics, Universitas Dhyana Pura, Bali, Indonesia. His research interests include computer security and reliability, computer networks, IoT, computing in mathematics, natural science, engineering, and medicine, as well as computing in social science, arts, and humanities. He is actively involved in empowering local entrepreneurs in Bali to realize their true potential through the use of information technology. He can be contacted at email: putuwida@undhirabali.ac.id.

I Gde Dhika Widarnandana hold bachelor's and master's degrees in Psychology, from Universitas Udayana, Indonesia, and Universitas Gadjah Mada, Indonesia respectively. He is an educator in Industrial and Organizational Psychology at the Faculty of Health and Science, Universitas Dhyana Pura, Bali, Indonesia. His research interests include industrial organizational psychology, training, and development, as well as experiential learning, including team building and tourism-psychology activities. He is also actively involved in a community empowerment project for developing community-based tourism from a psychological perspective. He can be contacted at email: gdedhika@undhirabali.ac.id.