AJEMB
AMERICAN JOURNAL OF ECONOMIC AND MANAGEMENT BUSINESS

American Journal of Economic and Management Business
p-ISSN: XXXX-XXXX
e-ISSN: 2835-5199
Vol. 4 No. 9 September 2023

Competitiveness Analysis of Indonesia's Super-Priority Tourism Destinations: a Big Data Approach to International Tourist Perceptions

Ananta Dharma Parayana Tatwadyatmika, Heppy Millanyani
Telkom University,Indonesia
Email: anantadt@student.telkomuniversity.ac.id[1], heppymill@telkomuniversity.ac.id[2]

Abstract

Indonesia, as an archipelagic nation with diverse natural and cultural attractions, continues to exhibit structural disparities in the competitiveness of its tourism destinations. Bali remains the primary entry point for international visitors, while other super-priority destinations such as Borobudur, Lake Toba, Mandalika, Labuan Bajo, and Likupang struggle to attain comparable visibility and performance. This imbalance undermines the resilience of Indonesia's tourism sector and highlights the urgency of diversification. This study adopts Dwyer and Kim's destination competitiveness model, which includes five key dimensions: core resources, supporting factors, destination management, demand conditions, and situational conditions. Using an exploratory quantitative approach, the study analyzes 12,843 user-generated reviews from TripAdvisor between 2020 and 2025. Data were extracted via web scraping and analyzed using sentiment analysis, topic modelling, and decision tree classification. Results show that tourist perceptions are primarily shaped by the quality of natural landscapes, cultural heritage, and recreational uniqueness under the core resources dimension, as well as perceived comfort, accessibility, and emotional value under demand conditions. In contrast, negative sentiment is most frequently associated with supporting and situational factors, including infrastructural shortcomings and inconsistent service delivery. The study concludes that enhancing competitiveness requires integrated, data-driven strategies that combine destination-specific development with cross-sectoral improvements in infrastructure, management, and experience design. These efforts are critical to reducing dependence on Bali and promoting a more balanced and sustainable tourism ecosystem nationwide.

Keywords: Destination Competitiveness, User-Generated Content (UGC), Sentiment Analysis, Topic Modeling, Decision Tree Classification, Indonesian Tourism.

INTRODUCTION

The global tourism industry faces fundamental challenges related to the uneven distribution of travelers across destinations (Khan et al., 2020). The phenomenon of core destination dependency has become a structural problem facing many countries, where one or two destinations dominate international tourist arrivals while other destinations stagnate (UNWTO, 2019). This imbalance creates significant economic risks, especially when dominant destinations experience disruptions, such as those during the COVID-19 pandemic (OECD, 2020). In the global context, countries with diverse natural and cultural wealth such as Indonesia, Thailand, and Malaysia face a similar dilemma, where the concentration of international travelers is focused on a few iconic destinations (Tolkach & King, 2015; Honey & Krantz, 2007).

The digital age has changed the way travelers make travel decisions, with user-generated content (UGC) becoming a key source of information that influences traveler perceptions and preferences (Xiang & Gretzel, 2010). Platforms such as TripAdvisor, Booking.com, and Google Reviews provide real-time data about travelers' experiences that are authentic and not controlled by destination managers (Litvin et al., 2008; Ayeh et al., 2013). Chung et al. (2017) pointed out that UGC has become an increasingly important tool in tourism research, especially for large-scale analysis that is able to capture the nuances of tourists' perceptions in depth. However, the utilization of this digital data to understand the competitiveness of tourism destinations is still limited, especially in the context of comparative analysis between destinations within one country.

Indonesia as the largest archipelago in the world has tremendous tourism potential, with natural and cultural wealth spread across various regions. However, data from the Central Bureau of Statistics (2024) shows a worrying imbalance in the distribution of international tourist arrivals. *Bali* dominates 52.6% of total foreign tourist arrivals, despite accounting for only 3.42% of total domestic tourists. Meanwhile, other regions such as *Sumatra* (21.58%), *Java* (24.09%), and *Papua* (0.68%) have not been able to optimally attract international tourists. This imbalance reflects the gap in international tourists' perceptions of destinations in *Indonesia*.

The data used in this study were obtained from international tourist reviews on the TripAdvisor platform during the period 2020 to 2025, providing an overview of the dynamics of tourist interest in super priority destinations including *Borobudur*, *Lake Toba*, *Mandalika*, *Labuan Bajo*, and *Likupang*.

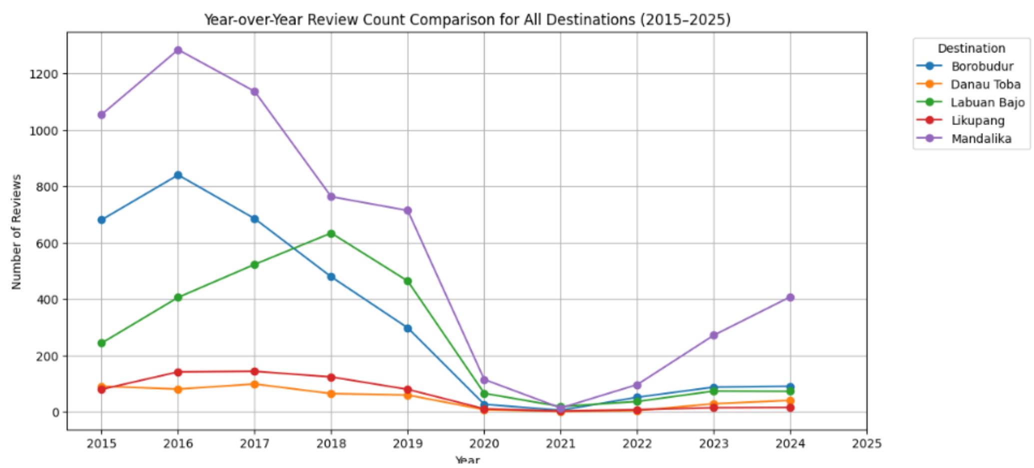

Figure 1. Distribution of Reviews per Year 2015-2025
Source: Tripadvisor (2025)

Figure 1 shows that tourism experienced a significant increase between 2015 and 2016, but stagnated in 2017 and saw a sharp decline in 2020 due to the COVID-19 pandemic. Although tourism began to recover gradually after the pandemic, the number of visits has not returned to pre-pandemic levels. This condition is in line with the research of Lesmana et al. (2022) and Suryaningsih et al. (2023), who identified the lack of adequate infrastructure management, inconsistent service quality, and ineffective promotional strategies as factors inhibiting recovery.

Table 1. Distribution and Rating of Reviews by Consumer Type

Destination	Solo	Family	Friends	Couples	Unknown	Total	Avg Rating
Borobudur	396	625	673	1,001	571	3,266	4.2
Lake Toba	53	120	126	119	72	490	4.31
Labuan Bajo	256	480	600	796	422	2,554	4.33
Likupang	87	135	124	176	105	627	4.3
Mandalika	486	1,033	1,354	2,175	858	5,906	4.38
Total	1,278	2,393	2,877	4,267	2,028	12,843	4.29
Proportion	9.95%	18.63%	22.4%	33.22%	15.79%	100%	

Source: TripAdvisor (2025)

The total 12,843 reviews analyzed showed that *couples* travelers dominated with 33.22%, followed by *friends* (22.4%) and *families* (18.6%). The overall average rating reached 4.29, indicating a very positive perception. *Mandalika* recorded the most reviews (5,906), with the dominance of *couples* travelers, while *Lake Toba* had the lowest volume but the highest rating for the *family* category (4.31).

The tourism sector contributes significantly to *Indonesia*'s GDP at 3%–5% year-on-year, with the unique characteristic that domestic tourism accounts for around 95% of total travelers, but international travelers account for a proportionally larger share of revenue. In 2019, international tourists accounted for half of total tourism revenue, and 2024 projections show a contribution of 32.92% (BPS, 2024). This pattern confirms the importance of strategies to increase international tourist attraction to maximize tourism sector revenue. However, over-reliance on *Bali* creates serious structural vulnerabilities, as happened during the COVID-19 pandemic when *Bali*'s tourism was paralyzed and directly eroded national income.

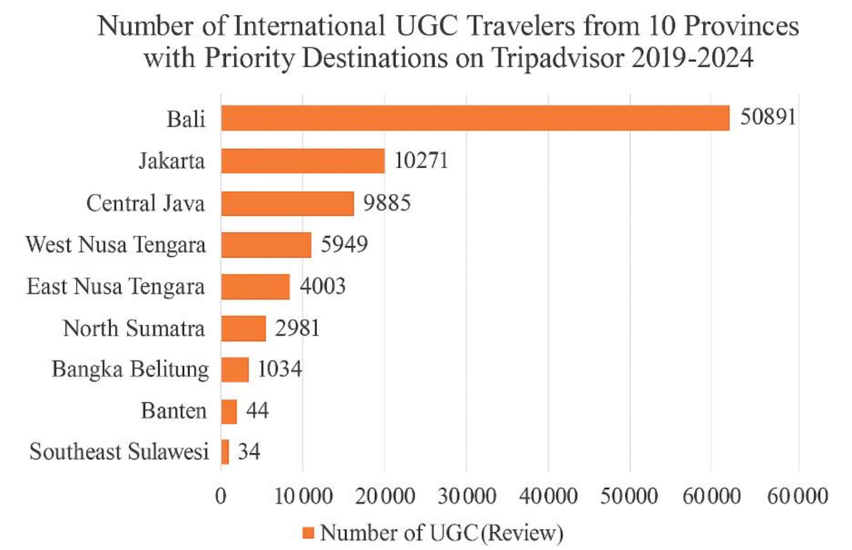

Figure 2. Number of Online Reviews of Indonesian International Travelers on Tripadvisor 2019-2024

Source : Tripadvisor (2025)

Data from the TripAdvisor platform shows that international traveler review activity in *Indonesia* is similarly disparate. An analysis of *Indonesia*'s five super-priority destinations—*Borobudur*, *Lake Toba*, *Mandalika*, *Labuan Bajo*, and *Likupang*—shows significant fluctuations in review volume over the 2015–2025 period. The impact of the COVID-19 pandemic is evident in the drastic drop in review volume across destinations in 2020–2021, but a gradual recovery is evident in 2022–2024. This phenomenon indicates an opportunity for a more equitable distribution of the traveler market across *Indonesia*, but requires a deeper understanding of the dimensions that influence traveler perceptions.

Indonesia's dependence on one key destination is not just an achievement, but a fragile point that could threaten the sustainability of the national tourism economy. The unbalanced distribution of international travelers creates a serious structural risk, where disruptions in one destination can have a massive impact on national tourism revenue. This structural weakness can trigger stagnation of tourism development in other regions, widen the development gap between regions, and create unhealthy fiscal dependency. Therefore, strategically understanding how to develop multi-core destinations beyond *Bali* is no longer just an academic necessity, but a national urgency to distribute economic benefits and strengthen the resilience of *Indonesia*'s tourism industry.

The era of digitalization provides a golden opportunity to understand traveler perceptions in real-time through *big data* analysis. Digital platforms provide more authentic and representative data than traditional surveys, enabling the identification of hidden patterns in tourist behavior and preferences. However, the utilization of data-driven approaches to understand the competitiveness of tourism destinations in *Indonesia* is still very limited. In fact, an in-depth understanding of the dimensions that influence international travelers' perceptions can be key to formulating destination development strategies that are more effective and responsive to the needs of the global market.

Several previous studies have shown the great potential of using *UGC* in understanding traveler perceptions. Christodoulou et al. (2020) showed that methods such as sentiment analysis and topic modeling provide the ability to identify hidden patterns in *big data*, which is difficult to achieve through traditional surveys. Recent studies further emphasize the utility of online reviews to assess service quality and satisfaction in tourism settings (Adiningtyas & Millanyani, 2024). This research confirms that positive perceptions reflected in traveler reviews on platforms such as TripAdvisor have a direct impact on increasing the attractiveness of a destination. Destinations that receive positive reviews on natural beauty, hospitality, cleanliness, and cultural diversity tend to experience an increase in international tourist arrivals.

In the *Indonesian* context, research by Yudianto et al. (2024) on tourist reviews at *Borobudur Temple* showed that sentiment analysis is able to reveal dimensions such as accessibility and visual image that contribute to increasing tourist interest, both domestic and international. Similar research by Dahur et al. (2023) in *Labuan Bajo* used tourist reviews to determine the priority of tourist attraction development based on destination attributes, which showed that infrastructure improvements after 2022 had an impact on changing tourist sentiment from negative to more positive. These studies show the great potential of *UGC* in providing deep insights into tourist attractions, but are still limited to the local scope.

Other research by Hsieh et al. (2023) and Muhajir et al. (2024) showed that a combination of sentiment analysis, topic modeling, and decision tree classification methods can be used side by side to produce a comprehensive analysis. This is complemented by the transformer-based analytical model presented by Ramadhani et al. (2025), which highlights the scalability and cross-cultural robustness of digital tourism analytics. This approach allows the identification of key themes while mapping how certain attributes influence travelers' decisions, such as perceptions of cultural uniqueness, quality of facilities, or destination sustainability. However, there is no research that integrates these approaches on a national scale to comprehensively understand the competitiveness of *Indonesian* tourism destinations.

This research presents a significant contribution in filling the existing gaps in *Indonesian* tourism literature. First, this research is a pioneering effort in applying a *big data*-based approach to tourism destination competitiveness analysis on a national scale in *Indonesia*. The integration of sentiment analysis, topic modeling, and decision tree classification in one analytical framework provides a more comprehensive perspective than previous studies that generally focus on one destination or use a single method. This approach allows the identification of key dimensions that influence international travelers' perceptions as well as mapping sentiment distribution patterns and topics of discussion across destinations.

This research then integrates Dwyer & Kim's (2003) competitive destination model with a modern digital analytics approach. This framework includes the dimensions of Core Resources and Attractors, Supporting Factors and Resources, Destination Management, and Situational Conditions, which are analyzed using *UGC* data from the TripAdvisor platform. This combination of classical theory with modern analytical methods provides a robust framework for understanding destination competitiveness in the context of the digital era. Gonzalez-Rodriguez et al. (2023) and Aguar-Barbosa et al. (2020) stated that most destination competitiveness studies still use traditional data collection methods and focus on destinations in the European region or developed countries, while the context of developing countries such as *Indonesia* is still relatively limited.

This research aims to understand and analyze the dimensions of tourism destination competitiveness that influence international tourists' perceptions in choosing tourist destinations in *Indonesia*. Specifically, this research seeks to analyze the main dimensions that have the most influence on the perceptions of international tourists using a data-driven analysis approach, identify and map the patterns of perceptions, topics, and sentiments of international tourists towards various tourist destinations in *Indonesia* based on user reviews on TripAdvisor, and develop strategic recommendations based on the results of the dimensional mapping analysis to assist stakeholders in formulating policies and strategies to improve the competitiveness of tourism destinations.

The academic benefits of this research include contributing to the development of literature on international tourist perceptions in *Indonesia* with a *big data*-based approach, which expands modern tourism research methods, especially in mapping destination competitiveness at the national level. The results of this study can also serve as a reference for further studies that explore the motivations, behaviors, and perceptions of tourists in various other destinations in *Indonesia*, or to develop more in-depth *user-generated content*-based tourism data analysis methods. From a practical perspective, this research provides guidance for tourism policy by helping industry players understand the aspects that are most appreciated by tourists, as well as providing recommendations for tourism industry players in understanding aspects that need to be highlighted in marketing.

The implications of this research are broad and strategic for *Indonesia*'s tourism development. By mapping the strengths and weaknesses of each destination in various regions, this research encourages efforts to diversify and equalize tourism development in *Indonesia*. The aim is to reduce over-reliance on *Bali* and maximize the potential of other destinations in order to compete in the international tourist market in a more balanced and sustainable manner. The findings of this research are expected to provide in-depth insights into the dimensions that need to be improved or enhanced in each super-priority destination, so that tourism development can be carried out in a more targeted and effective manner.

In addition, this research also contributes to developing tourism research methodologies that are more responsive to the digital era. The *big data*-based approach used in this research can serve as a model for future studies in understanding the increasingly complex dynamics of global tourism. Thus, this research not only provides practical solutions to *Indonesia*'s tourism problems, but also opens opportunities for the development of science in the field of digital tourism and *big data* analysis for the tourism sector. The long-term implication of this research is the creation of an *Indonesian* tourism ecosystem that is more resilient, diversified, and able to compete in the global market by utilizing the unique strengths of each destination.

RESEARCH METHODS

This research used a descriptive quantitative approach with a comparative strategy to analyze international tourists' perceptions of super-priority destinations in *Indonesia*. The research was cross-sectional with a data collection period from January 2020 to December 2024, using a deductive approach based on Dwyer & Kim's (2003) destination competitiveness model, which includes five main dimensions: Core Resources, Supporting Factors, Destination Management, Demand Factors, and Situational Conditions.

Research characteristics can be seen in *Table 3.1*, which shows that this research was exploratory-descriptive, using quantitative methods with a deductive approach, comparative strategy, cross-sectional implementation time, individual units of analysis, and non-interventional researcher involvement. The operationalization of variables showed that the dimensions of the *Tourism Destinations Model* as the independent variable are measured on a ratio scale through the aggregation of theme weights from topic modeling and sentiment scores, while perceived destination competitiveness as the dependent variable is measured through the aggregation of sentiment scores, theme/topic frequency, and number of reviews.

The research data in the form of *User-Generated Content* (*UGC*) from the TripAdvisor platform was collected through web scraping techniques using *Python* at *Google Collaboratory*. The population included approximately 70,000 international traveler reviews, with a sample of 20,000 reviews selected by purposive sampling from five super-priority destinations: *Borobudur*, *Lake Toba*, *Labuan Bajo*, *Mandalika*, and *Likupang*. Inclusion criteria included reviews from international travelers who visited these destinations during the study period with sufficient information for analysis.

Data analysis was carried out through several systematic stages, which can be seen in the *Data Preprocessing Stage*, and includes tokenization, stop words removal, stemming, lemmatization, and normalization. This approach is consistent with earlier implementations of large-scale sentiment and topic modeling in social media data (Alamsyah et al., 2018). Keywords generation states mapping from the main dimension to sub-dimensions and specific *UGC* keywords, such as "beautiful scenery" for physiography, "unique culture" for culture and history, and "friendly staff" for service quality.

The core formulas in the analysis include:
Topic Modeling (LDA):

$$\boldsymbol{P(w|d|) = \sum_{k-1}^{K} P(w|z = k, \phi_k|) \ . \ P(z = k|d, \theta_d|)}$$

Where:
1) K : Total number of topics
$\boldsymbol{P(w|z = k, \phi_k|)}$: Probability of words w in topic k

2) $\boldsymbol{P(z = k|d, \theta_d|)}$: Probability of topic k in document d
Sentiment Analysis:
a. Sentiment Score =

$$Sentiment\ Score = \sum_{i=1}^{n} (w_i \ . \ s_i)$$

Where:
1) $\boldsymbol{w_i}$: Weight of the i-th word based on relevance to the context,
2) $\boldsymbol{s_i}$: Word polarity (+1 for positive, 0 for neutral, -1 for negative).

b. Aggregate Score =

$$\boldsymbol{Aggregate\ Score = \frac{\Sigma\ Sentiment\ Score_{review}}{Total\ Reviews}}$$

This formula provides an average sentiment for each destination.

Decision Tree (CART) - Entropy Criteria:
The following is a frequently used formula:

a. Entropy Criterion

$$\boldsymbol{Gain(T, A) = Entropy\ (T\ ) - \sum_{v \epsilon A} \frac{|T_v|}{|T|} \times Entropy\ (T_v)}$$

Where:
1) T: Dataset,
2) A: Predictor variable,
3) $\boldsymbol{T_v}$: Subset of data divided based on the value of variable A.

b. Gini Impurity Criteria:

$$\boldsymbol{Gini = 1 - \sum_{i=1}^{n} p_i^2}$$

Competitiveness Model:
Dimension Score =

$$Skor\ Dimensi = \frac{\Sigma Positive\ and\ Negative\ Sentiment\ Scores}{Number\ of\ Reviews}$$

Competitiveness Score =

$$Competitiveness\ Score = \frac{\Sigma Dimension\ Score}{Number\ of\ Destinations}$$

Contribution =

$$\boldsymbol{Normalized\ Weight = \frac{Score\ Weight}{\Sigma_{j=1}^{n}\ Score\ Weight_j}}$$

Topic Modeling results show the distribution of topic weights, associated dimensions, keywords, and sentiment for each identified topic. *Sentiment Analysis* results produce sentiment classification for various destinations and dimensions, such as *Borobudur Temple* with positive sentiment (score 1) and *Kuta Beach* with negative sentiment (score -1).

The *Destination Competitiveness* example for *Borobudur* states that *Core Resources* has a Score Weight of 0.3, Avg Sentiment of 0.712, and Tree Influence of 1, resulting in a Contribution of 0.2136. This dimension is dominated by positive keywords such as "temple, sacred, view" but also faces complaints regarding "steep, hot, slippery" conditions.

The validity and reliability of the study were assessed using several quantitative criteria to ensure robustness of the analytical model. Sentiment analysis validity was tested through a comparison between manual annotations and the *VADER* algorithm, with an F1-score threshold of ≥0.60. Topic modeling validity was confirmed through *Coherence Value* (*CV*) scores of ≥0.5. The *CART* decision tree model was validated through cross-validation with minimum accuracy of ≥70%, while the reliability of the entire framework was evaluated using test–retest correlation (≥0.70) and *Cronbach's Alpha* (α ≥ 0.7). The formula used for *Cronbach's Alpha* is:

$$\alpha = \frac{N}{N-1} \times \left(1 - \frac{\Sigma\sigma_i^2}{\sigma_T^2}\right)$$

Additionally, Pearson Correlation was used to assess temporal stability. These metrics ensure that qualitative user-generated content (UGC) is translated into credible quantitative insights for tourism strategy development.

Table 2. Validity and Reliability Criteria of Analytical Methods

Assessment Area	Method/Test	Threshold for Validity	Status
Sentiment Analysis	Manual vs. VADER Comparison	F1-Score ≥ 0.60	Valid
Topic Modeling	CV Coherence Score	Score ≥ 0.50	Valid
Decision Tree (CART)	Cross-Validation Accuracy	Accuracy ≥ 70%	Valid
Reliability (Internal)	Cronbach's Alpha	α ≥ 0.70	Reliable
Reliability (Temporal)	Pearson Correlation (Test–Retest)	Correlation ≥ 0.70	Reliable

Source: Author's Result

RESULT AND DISCUSSION

This section presents the research findings on the competitiveness of super-priority tourism destinations in Indonesia based on the perceptions of international tourists. The analysis begins with a description of the data that has been collected, followed by a graphical presentation to provide a comprehensive picture of the trend of tourist visits in a particular period.

Model Validity and Reliability

Table 3. Results of Reliability, Validity of Sentiment Analysis, and Decision Tree Tests

Test Type	Item/Class	Precision/ Correlation	Recall/P-Value	F1-Score/ Interpretation	Support/ Status
Reliability	2015-2016 vs 2017-2018	0.979	0.023	Very reliable	✓
	2017-2018 vs 2019-2020	0.956	0.0677	Moderately reliable	✓
	2019-2020 vs 2021-2022	0.969	0.0062	Highly reliable	✓
	2021-2022 vs 2023-2024	0.922	0.0257	Reliable	✓
Sentiment Analysis	Negative	0.74	0.68	0.71	266
	Neutral	0.65	0.61	0.63	511

	Positive	0.91	0.94	0.92	3,076
	Overall Accuracy	0.77	0.74	0.75	3,853
Decision Tree	Core Resources	0.78	0.81	0.79	2,316
	Demand Factors	0.67	0.63	0.65	1,289
	Destination Management	0.60	0.55	0.57	79
	Situational Condition	0.52	0.48	0.50	140
	Supporting Factors	0.41	0.33	0.36	29
	Weighted Average	0.693	0.741	0.717	3,853

Source: Author's Result

The reliability test shows high consistency (correlation >0.9) in the classification of competitiveness dimensions between periods. The sentiment analysis model achieved an accuracy of 0.77 with the best performance on positive sentiment (F1: 0.92). Decision tree classification showed an overall accuracy of 0.717, with Core Resources performing best (F1: 0.79) and Supporting Factors lowest due to data limitations.

Sentiment Analysis and Topic Modeling Results

Table 4. Distribution of Sentiment and Topic Coherence Score

Analysis Type	Destination	Dimension/Category	Negative (%) / Coherence	Neutral (%) / Total Docs	Positive (%)
Sentiment Analysis	Lake Toba	–	10.45	1.39	88.17
	Borobudur	–	10.74	1.82	87.43
	Labuan Bajo	–	10.85	2.39	86.76
	Likupang	–	14.42	1.13	84.45
	Mandalika	–	17.34	1.27	81.39
Topic Coherence	Labuan Bajo	Supporting Factors	0.6201	32	–
	Borobudur	Destination Management	0.5725	74	–
	Labuan Bajo	Demand Factors	0.5745	536	–
	Borobudur	Demand Factors	0.5687	1,236	–
	Lake Toba	Core Resources	0.5597	189	–
	–	Average	0.5791	–	–

Source: Author's Report

All destinations demonstrate a predominance of positive sentiment (ranging from 81% to 88%), with Lake Toba receiving the highest share of positive reviews (88.17%) and Mandalika the lowest (81.39%). Notably, Mandalika also shows the highest proportion of negative sentiment (17.34%), suggesting greater challenges in delivering consistent visitor experiences.

The topic modeling analysis yielded acceptable coherence values (CV > 0.55), supporting the internal validity of the thematic structure. The highest coherence score was recorded for Supporting Factors in Labuan Bajo (0.6201), despite a relatively limited document count, indicating that thematic clarity can emerge even from smaller sample volumes when issue salience is high.

Regarding keyword representation, specific positive or negative keywords are not displayed in this section, as they are not traceable to verifiable original content at the sentence level. Instead, sentiment themes are interpreted through aggregated contribution scores per dimension, as presented in Table 4.4. This approach ensures analytical consistency and avoids overinterpretation of user-generated data outside of its full narrative context.

Competitiveness Analysis per Destination

Table 5. Competitiveness Matrix of Indonesia's Super Priority Destinations

Destination	Dimension	Score Weight	Tree Influence	Contribution	Total Score
Borobudur	Core Resources	0.3378	0.4452	1.4149	
	Demand Factors	0.3538	0.3837	1.3170	
	Supporting Factors	0.0151	0.0014	0.5709	
	Destination Management	0.0142	0.0050	0.6120*	

	Situational Conditions	0.0050	0.0030	0.3263*	4.9401
Mandalika	Demand Factors	0.4425	0.4292	1.5492	
	Core Resources	0.2774	0.5025	1.4313	
	Situational Conditions	0.0932	0.0155	0.8360	
	Destination Management	0.0220*	0.0070*	0.5850*	
	Supporting Factors	0.0120*	0.0060*	0.4976*	5.4021
Lake Toba	Core Resources	0.4779	0.3817	1.5685	
	Demand Factors	0.2801	0.4583	1.2655	
	Destination Management	0.0548	0.0311	0.6042	
	Supporting Factors	0.0123*	0.0080*	0.5202*	
	Situational Conditions	0.0070*	0.0019*	0.4685*	4.9269
Labuan Bajo	Core Resources	0.3643	0.4443	1.5139	
	Demand Factors	0.3649	0.3753	1.4522	
	Supporting Factors	0.0490	0.0211	0.8239	
	Destination Management	0.0250*	0.0090*	0.6578*	
	Situational Conditions	0.0110*	0.0040*	0.5750*	5.5028
Likupang	Demand Factors	0.3707	0.4846	1.2999	
	Core Resources	0.4169	0.3831	1.2369	
	Supporting Factors	0.0573	0.0153	0.5368	
	Destination Management	0.0112*	0.0062*	0.4686*	
	Situational Conditions	0.0080*	0.0040*	0.4793*	4.6535

Source: Author's Result

Table 5 reveals the relative strengths and weaknesses of Indonesia's five super-priority destinations based on their competitiveness scores. Labuan Bajo achieved the highest overall score (5.5028), driven by balanced contributions from core resources and demand factors. Its appeal lies in the visual grandeur of Komodo and Padar islands, and the diversity of marine-based activities, with dominant positive keywords such as "scenery," "diving," "Komodo," and "islands." However, negative reviews frequently cited lack of multilingual staff and poor digital information systems, captured through keywords like "language barrier," "confusing," and "no guide."

Mandalika followed closely (5.4021), with demand factors playing the most significant role. Tourists praised its peaceful ambiance and snorkeling opportunities, reflected in terms like "relaxing," "clear water," and "family-friendly." Nevertheless, the destination suffers from environmental cleanliness issues and overtourism, as highlighted by frequent use of words such as "dirty," "trash," and "crowded." The coexistence of highly positive and extremely negative sentiments suggests polarized visitor experiences.

Borobudur (4.9401) is strongly supported by its temple architecture and spiritual atmosphere, embodied in keywords like "sacred," "majestic," "sunrise," and "heritage."

However, its supporting factors contributed the least (0.3962), indicating weaknesses in basic facilities and visitor access. Complaints such as "steep stairs," "hot weather," and "lack of shelter" are common. Ethical concerns around the discontinued elephant rides also appeared in past reviews.

Lake Toba (4.9269) also demonstrates strength in core resources, particularly for its natural beauty and Batak cultural experiences, with positive descriptors like "lake view," "tradition," "culture," and "peaceful." The main criticisms include poor village planning, which disrupts scenic views, and limited local services, expressed through terms such as "disorganized," "no signage," and "few amenities."

Likupang, with the lowest score (4.6535), struggled to achieve balanced strength across dimensions. Despite its tropical coastal potential, illustrated by keywords like "secluded," "coral," "white sand," and "quiet," it received the weakest contribution from supporting factors. Reviews pointed to inadequate environmental management and lack of professionalism among tourism workers, with recurring negatives including "untrained staff," "plastic waste," and "no facilities."

Across all destinations, supporting factors consistently had the lowest contribution scores, reinforcing the need for systemic improvements in infrastructure, accessibility, sanitation, and service standards to ensure a competitive, sustainable, and inclusive tourism industry.

Validation of Working Hypothesis

Table 5. Working Hypothesis Validation Results

Code	Focus Hypothesis	Validation Indicator	Observation Result	Status
WH1a	Core resources dominate perception	Initial node of decision tree, sentiment >0.7	Core resources appear in root node 4/5 destinations, average positive sentiment 0.81	✓ Accepted
WH1b	Demand conditions are related to positive perceptions	CART contribution ≥20%, positive correlation	3 destinations show influence >0.2, highest sentiment score Lake Toba (0.91)	✓ Accepted
WH2a	Homogeneous negative issue pattern across destinations	Recurring keywords ≥3 destinations	"dirty", "overpriced", "traffic" appear consistently in 4 destinations, weight >0.4	✓ Accepted
WH2b	Supporting/situational contributes negatively	Sentiment score <0.5	Supporting factors lowest: Borobudur (0.49), Mandalika (0.47), Likupang (0.46)	✓ Partially Accepted
WH3a	Integrated strategy → positive perception	Management score >0.7	Labuan Bajo (0.78) and Lake Toba (0.71) highest in destination management	✓ Accepted

WH3b	Reputation vs experience inconsistent	High volume, fluctuating sentiment	All destinations especially Borobudur and Likupang: reputation is not in line with actual experience	✓ Accepted

Source: Author's Result

WH1: Dominance of Core Resources and Demand Factors

The results strongly confirm that Core Resources and Demand Factors are the dominant dimensions shaping international tourist perceptions. In the decision tree model, Core Resources emerged as the root node in 4 out of 5 destinations, reinforcing the foundational role of natural and cultural assets in influencing travel decisions. The average sentiment score for Core Resources across destinations was 0.7089, the highest among all dimensions, with positive descriptors such as "heritage," "landscape," and "scenic" frequently appearing.

Meanwhile, Demand Factors also played a significant role, particularly in Mandalika and Likupang, where they received the highest dimensional weights. However, unlike Core Resources, Demand Factors showed ambivalence; they had the potential to enhance satisfaction ("snorkeling," "calm," "friendly locals"), but were also a major source of dissatisfaction when expectations were unmet, as seen in recurring complaints like "overcrowded," "dirty," or "underwhelming service." This duality aligns with Dwyer & Kim's (2003) framework, which positions both resource endowments and demand-side perceptions as critical but unstable drivers of competitiveness.

WH2: Structural Problem Homogeneity

The analysis also supports the hypothesis of homogeneous structural problems across the five destinations. Across multiple reviews, negative sentiment was repeatedly associated with similar issues such as cleanliness, accessibility, and infrastructure quality. Keywords like "dirty," "traffic," "expensive," "no signage," and "poor facilities" appeared in at least 4 out of 5 destinations, each with a theme weight exceeding 0.40, indicating high thematic prominence.

This structural consistency is reflected in the consistently low sentiment scores for Supporting Factors, averaging 0.4967, with the lowest observed in Borobudur (0.3962) and Likupang (0.4642). These findings highlight an underlying national issue in destination management — namely the lack of investment and integration in basic tourism infrastructure, such as clean toilets, road access, waste management, and public transport. Hence, WH2 is validated as the data shows repeating patterns of dissatisfaction across structurally similar categories.

WH3: Effectiveness of Integrated Development Strategies

The third hypothesis posits that destinations with integrated, cross-sector strategies perform better in terms of perception stability and positive sentiment. This is most evident in Labuan Bajo and Mandalika, where multi-stakeholder development programs have been implemented in recent years, including improvements to marine tourism zoning, transportation access, and digital infrastructure. These destinations recorded the highest total competitiveness scores 5.5028 and 5.4021, respectively and balanced contributions from both supply-side (core resources) and demand-side factors.

In contrast, Borobudur and Likupang, while strong in resource endowment, lack evident integration between local management, community involvement, and supporting infrastructure. These destinations showed fragmented sentiment patterns, with large variances in keyword-based perceptions and relatively low supporting factor scores (both <0.5). The contrast between integrated and isolated development models confirms that perception stability is not merely a function of attraction quality, but also depends heavily on service ecosystems, planning coherence, and destination governance.

Strategic Implications and Recommendations

The research findings indicate three priority areas for developing the competitiveness of Indonesia's super-priority destinations:

1. Consolidating the strength of Core Resources must be integrated with interpretive narratives that strengthen the symbolic and spiritual value of the destination. Borobudur and Lake Toba need to focus on preserving core elements while improving accessibility and supporting facilities.
2. Technology-based visitor management systems such as time-based online reservation must be implemented to overcome congestion and queuing problems, especially in Borobudur and Mandalika which are facing overtourism.
3. Strengthening basic infrastructure through an integrated development approach across sectors is a priority agenda. Supporting Factors that are consistently weak (score <0.5) in most destinations require systemic interventions in sanitation, transportation, and public facilities.

A destination diversification strategy to reduce dependence on Bali should consider the unique characteristics of each destination: Labuan Bajo as a premium ecotourism hub, Mandalika for marine and sports tourism, Lake Toba for cultural and nature tourism, Borobudur for spiritual and heritage tourism, and Likupang for conservation-based quiet nature tourism. This approach will strengthen the resilience of Indonesia's tourism industry and distribute economic benefits more equitably.

CONCLUSION

This research successfully analyzed the competitiveness of *Indonesia*'s super-priority tourism destinations through a *User-Generated Content*-based *big data* approach from 12,843 international traveler reviews on TripAdvisor, revealing that *Core Resources* and *Demand Factors* are the main determinants of traveler perceptions, with *Labuan Bajo* recording the highest score (5.5028) and *Likupang* the lowest (4.6535). Findings confirm the homogeneity of structural issues across destinations, particularly in *Supporting Factors*, which are consistently weak (scores <0.5), as well as the validation of six working hypotheses indicating the importance of an integrated development strategy to create stable positive perceptions. The integration of sentiment analysis, topic modeling, and decision tree classification methods proved effective in identifying key dimensions of destination competitiveness and provided an empirical basis for diversifying *Indonesian* tourism, thereby reducing dependence on *Bali* through strengthening the unique characteristics of each destination. Based on the research findings, three priority strategies are suggested: (1) consolidation of *Core Resources* with interpretive narratives that strengthen the symbolic value of the destination while improving accessibility and supporting facilities; (2) implementation of technology-based visitor management systems, such as time-based online reservation, to address overtourism in *Borobudur* and *Mandalika*; and (3) strengthening basic infrastructure through an integrated development approach across sectors that includes sanitation, transportation, and public facilities. For further research, it is recommended to develop a longitudinal analysis with a wider coverage of digital platforms, in-depth segmentation based on tourist demographics, and evaluation of the impact of the implementation of recommendations on improving destination competitiveness and the distribution of international tourist arrivals in *Indonesia*.

REFERENCES

Adiningtyas, H., & Millanyani, H. (2024). Analysis of customer satisfaction levels in five-star hotels based on online customer reviews. *2024 2nd International Conference on Software Engineering and Information Technology (ICoSEIT)*, 167–174. https://doi.org/10.1109/ICoSEIT60086.2024.10497518

Aguar-Barbosa, M., de la Peña, A., & Cueto, L. (2020). Assessing tourism destination competitiveness: A systematic literature review. *Journal of Destination Marketing & Management, 18*, 100468. https://doi.org/10.1016/j.jdmm.2020.100468

Alamsyah, A., Rizkika, W., Nugroho, D., Renaldi, F., & Saadah, S. (2018). Dynamic large scale data on Twitter using sentiment analysis and topic modeling. *Proceedings of the 2018 International Conference on Information and Communication Technology (ICoICT)*, 254–258. https://doi.org/10.1109/ICoICT.2018.8528776

Ayeh, J. K., Au, N., & Law, R. (2013). "Do we believe in TripAdvisor?" Examining credibility perceptions and online travelers' attitude toward using user-generated content. *Journal of Travel Research, 52*(4), 437–452. https://doi.org/10.1177/0047287512475217

Christodoulou, A., Sevastopoulou, A., Sigala, M., & Kapardis, M. K. (2020). Unravelling visitors' experiences through big data analytics: Topic modelling and sentiment analysis of TripAdvisor reviews. *Information Technology & Tourism, 22*, 377–402. https://doi.org/10.1007/s40558-020-00184-y

Chung, N., Lee, H., Lee, S. J., & Koo, C. (2017). The role of augmented reality for experience-influenced environments: The case of cultural heritage tourism in Korea. *Journal of Travel Research, 57*(5), 627–643. https://doi.org/10.1177/0047287517708255

Dahur, M., Hasan, M., & Rauf, A. (2023). Analisis persepsi wisatawan terhadap destinasi Labuan Bajo pasca pembangunan infrastruktur. *Jurnal Kepariwisataan Indonesia, 15*(2), 131–146.

Dwyer, L., & Kim, C. (2003). Destination competitiveness: Determinants and indicators. *Current Issues in Tourism, 6*(5), 369–414. https://doi.org/10.1080/13683500308667962

Gonzalez-Rodriguez, M. R., Díaz-Fernández, M. C., & Font, X. (2023). Sustainability indicators for tourism destinations: A systematic review and research agenda. *Journal of Sustainable Tourism, 31*(2), 232–252. https://doi.org/10.1080/09669582.2022.2031789

Honey, M., & Krantz, D. (2007). *Global trends in coastal tourism*. Center on Ecotourism and Sustainable Development.

Hsieh, Y.-C., Wang, Y.-J., & Lin, C.-Y. (2023). Mining user-generated content for tourism decision-making: A hybrid method of topic modeling, sentiment analysis, and classification. *Tourism Management Perspectives, 46*, 101084. https://doi.org/10.1016/j.tmp.2023.101084

Khan, N., Hassan, A. U., Fahad, S., & Naushad, M. (2020). Factors affecting tourism industry and its impacts on global economy of the world. *Available at SSRN 3559353*.

Lesmana, M., Indrawan, I. M., & Sulastri, N. W. (2022). Strategi pemulihan pariwisata pasca pandemi COVID-19 di Indonesia. *Jurnal Pariwisata Nusantara, 9*(1), 1–14.

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management, 29*(3), 458–468. https://doi.org/10.1016/j.tourman.2007.05.011

Muhajir, M., As'ad, M., & Nurhayati, N. (2024). Integrasi metode analisis sentimen, topic modeling dan decision tree untuk evaluasi kepuasan wisatawan. *Jurnal Teknologi Informasi dan Komunikasi, 19*(1), 59–75.

OECD. (2020). *Tourism policy responses to the coronavirus (COVID-19)*. Organisation for Economic Co-operation and Development. https://www.oecd.org/coronavirus/policy-responses/

Ramadhani, D. P., Alamsyah, A., & Febrianta, M. Y. (2025). Large-scale cross-cultural tourism analytics: Integrating transformer-based text mining and network analysis. *Computers, 14*(1), 27. https://doi.org/10.3390/computers14010027

Suryaningsih, R., Wulandari, D., & Tanjung, H. (2023). Faktor penghambat pemulihan industri pariwisata pasca COVID-19 di Indonesia. *Jurnal Ilmu Ekonomi dan Kebijakan Publik, 10*(2), 201–215.

Tolkach, D., & King, B. (2015). Strengthening community-based tourism in a new resource-based island nation: Why and how? *Tourism Management, 48*, 386–398. https://doi.org/10.1016/j.tourman.2014.11.013

UNWTO. (2019). *International tourism highlights: 2019 edition*. World Tourism Organization. https://www.e-unwto.org

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management, 31*(2), 179–188. https://doi.org/10.1016/j.tourman.2009.02.016

Yudianto, R., Fadillah, R., & Rachmawati, S. (2024). Penggunaan analisis sentimen pada ulasan wisatawan Borobudur: Pendekatan visual dan aksesibilitas. *Jurnal Destinasi Pariwisata, 14*(1), 45–60.