Environment and Natural Resources Journal

Volume 24
Issue 6
2026

Identifying Geospatial Priority Hotspots of Tourism Sustainability Risk using a Destination Sustainability Index (DSI) Devised from Online Tourist Dissatisfaction and Getis-Ord G_i^* Statistics

Tobthong Chancharoen[1], Asamaporn Sitthi[2], Parinya Nakpathom[3], and Narong Pleerux[1*]

[1]*Faculty of Humanities and Social Sciences, Burapha University, Thailand*
[2]*Faculty of Social Sciences, Srinakharinwirot University, Thailand*
[3]*International College, Burapha University, Thailand*

ARTICLE INFO

Received: 27 Nov 2025
Received in revised: 26 Jun 2026
Accepted: 30 Jun 2026
Published online: 3 Sep 2026
DOI: 10.32526/ennrj/24/20250343

Keywords:
Destination Sustainability Index (DSI)/ Hotspot analysis (G_i^*)/ Sentiment analysis/ Spatial clustering/ Tourist dissatisfaction data

*** Corresponding author:**
E-mail: narong_p@go.buu.ac.th

ABSTRACT

Managing tourism sustainability effectively requires identifying specific localized risks, which remains a significant challenge for many destinations. This research aims to identify priority hotspots of unsustainability in marine and beach tourism destinations in Eastern Thailand by employing geospatial analysis and tourist-perceived data to support targeted policy decisions. Drawing from 11,047 geotagged Tripadvisor reviews as a primary data source, the methodology commenced with processing these reviews using a high-precision transformer-based (BERT) sentiment analysis model, achieving 87% precision in the negative class. This was followed by N-gram analysis to extract specific problem phrases, which were used to construct the Destination Sustainability Index (DSI) across three main dimensions: Environmental (DSIEnv), Congestion/Access (DSIAcc), and Service/Value (DSISrv). The DSI was then subjected to hotspot analysis (Getis-Ord Gi*) to confirm the statistical significance of spatial clustering. Results confirm that Environmental problems (rubbish/dirty) and Congestion/Access issues (crowded, boat) are the most frequent concerns. The analysis clearly demonstrated that tourism dissatisfaction is not randomly distributed but localized in statistically significant spatial locations. Priority hotspots showing overlapping risks in Chonburi Province require immediate intervention, while the successful management of the Koh Kood cluster offers a blueprint for sustainable development.

HIGHLIGHTS

This study develops the Destination Sustainability Index (DSI) using tourist dissatisfaction data from 11, 047negative online reviews. Hotspot analysis (G_i^*) was then employed to map spatial risks. Key findings confirm that the highest risk hotspots overlap in Congestion/Access and Service/Value dimensions within the Chonburi cluster (Koh Larn, Si Racha). These areas require urgent structural investment and the implementation of Carrying Capacity limits. Conversely, the Koh Kood cluster (Trat) is the most robust coldspot across multiple dimensions and should be established as the best practice model. The research demonstrates the effectiveness of using tourist dissatisfaction data for DSI and G_i^* in precise policy prioritization.

1. INTRODUCTION

As a cornerstone of Thailand's economy, tourism contributes approximately 17% to the national GDP (MOTS, 2023), with marine and beach attractions in the Eastern region remaining primary destinations (TAT, 2024). However, this tourism-driven economic growth has triggered rapid environmental deterioration, including increased waste generation, pollution, and congestion (MOTS, 2023), issues now exacerbated by global climate change (TAT, 2024). Implementing sustainable resource management is thus an urgent necessity aligned with Thailand's National Strategy (MOTS, 2023), requiring empirical evidence for targeted policy decisions.

In the digital era (Cui and He, 2019), TripAdvisor reviews (Jeacle and Carter, 2011) serve as high-volume Big Data reflecting authentic tourist perspectives (Li et al., 2017; Alfred et al., 2023). Integrating sentiment analysis of these reviews

Citation: Chancharoen T, Sitthi A, Nakpathom P, Pleerux N. Identifying geospatial priority hotspots of tourism sustainability risk using a Destination Sustainability Index (DSI) devised from online tourist dissatisfaction and Getis-Ord G_i^* statistics. Environ. Nat. Resour. J. 2026;24(6):689-700. (https://doi.org/10.32526/ennrj/24/20250343)

(Valdivia et al., 2017) with geospatial techniques (Jabalameli et al., 2022) enables a deeper understanding of destination dynamics. While most previous studies have focused primarily on positive indicators, (Singgalen et al., 2024; Charfaoui and Mussard, 2024) this approach often overlooks the granular, spatial manifestations of systemic failure (Gadekar, 2024). Therefore, a significant research gap exists in utilizing negative sentiments to map spatial problem density through advance language models like BERT (Yuan, 2024; Xi et al., 2024).

Consequently, this research employs N-gram analysis (Ojo et al., 2021) and a novel Destination Sustainability Index (DSI) derived from geo-locatable negative feedback to identify and map "Priority Hotspots." This methodology aims to categorize problems into Environmental, Congestion/Access, and Service/Value dimensions, providing empirical geospatial evidence for government agencies to prioritize resource restoration and management interventions.

2. METHODOLOGY

2.1 Study area

Focusing on the coastal and marine attractions of Eastern Thailand, this study encompasses the strategic provinces of Chonburi, Rayong, Chanthaburi, and Trat (Figure 1). This region serves as a pivotal economic engine for the national tourism sector, primarily driven by intensive maritime and beach-based activities (MOTS, 2023). However, the high volume of tourism has exerted substantial pressure on fragile ecosystems (Baltranaite et al., 2025), evidenced by resource degradation from uncontrolled tourism expansion (ONEP, 2024), critical marine waste challenges (Abirami, 2024), and arrivals that frequently exceed carrying capacity limits (Yusoh et al., 2021). Consequently, this area provides a representative case study for applying geospatial analysis to identify spatial problem density and assess marine resource degradation under high-intensity tourism pressure.

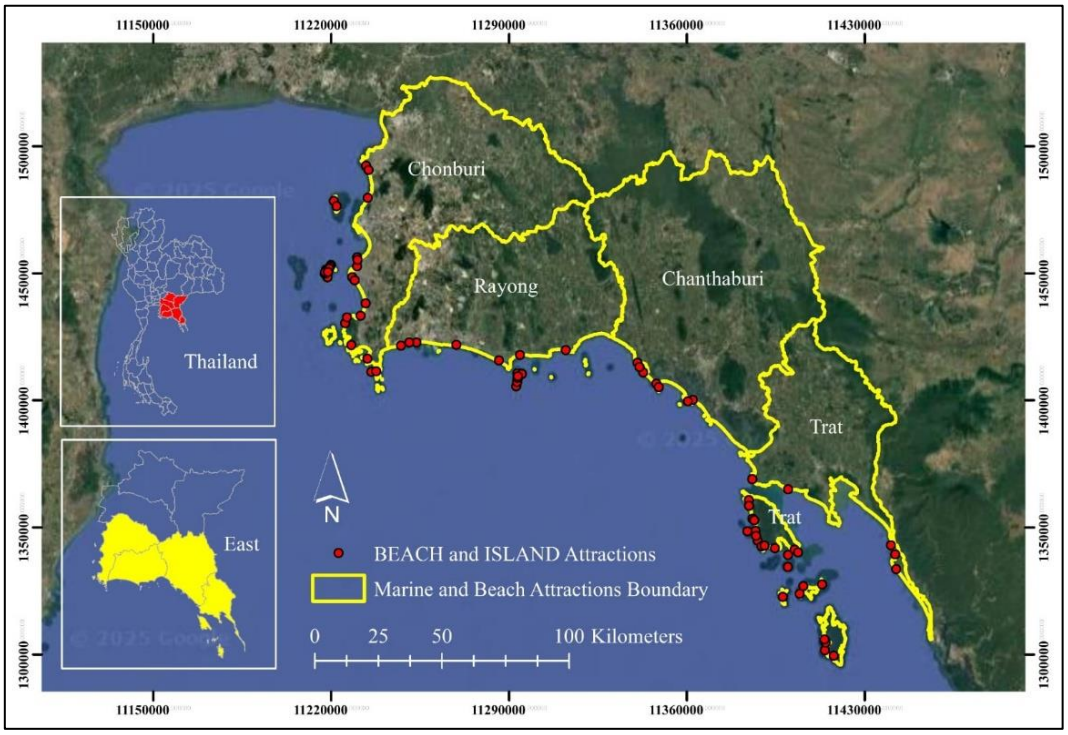


Figure 1. Study area and distribution of marine and beach attractions in Eastern Thailand

2.2 Data collection and preprocessing

2.2.1 Data source and scope

International tourist reviews of marine and beach attractions in Eastern Thailand were retrieved from TripAdvisor (Valdivia et al., 2017) spanning a ten-year period from 2014 to 2023. The dataset includes 11,047 reviews categorized as BEACH and ISLAND attractions (DOT, 2018). Data acquisition was performed through web scraping using Python, specifically leveraging the BeautifulSoup, Pandas, and Selenium libraries (Muthukadan, 2024).

2.2.2 Text cleaning and normalization

The raw review data underwent a systematic multi-step cleaning process to isolate essential features for sentiment analysis (Ramadhani et al., 2021). This included converting text to lowercase and removing punctuation, URLs, numbers, and special characters. Stop words were eliminated; however, negation terms such as "no" and "not" were retained to maintain the contextual integrity of negative sentiments. Finally, lemmatization was applied to reduce words to their lexical base forms before tokenizing the text into individual lexical unit (Chakriswaran et al., 2019).

2.3 Deep learning-based sentiment analysis and feature extraction

2.3.1 Deep learning model for sentiment classification

This study utilized the Pysentimiento framework (Pérez et al., 2024), which is built upon the BERT (Bidirectional Encoder Representations from Transformers) architecture (Devlin et al., 2019). BERT was selected for its bidirectional contextual learning architecture (Yuan, 2024), which allows the model to capture nuances in tourist language—such as sarcasm or complex negations—more effectively than traditional lexicon-based approaches VADER.

2.3.2 Model evaluation and validation

Performance was validated through hyperparameter fine-tuning and systematic error checks to mitigate classification bias (Devlin et al., 2019). Evaluation metrics derived from the confusion matrix (Table 1) included precision (P), recall (R), F1-score, and overall accuracy (A) (Pleerux et al., 2023). Specific focus was placed on the precision of the negative class to ensure the reliability of the dataset for subsequent geospatial analysis (Equation 1-4).

Table 1. Confusion matrix for sentiment model validation

Actual/prediction	Positive	Neutral	Negative
Positive	a	b	c
Neutral	d	e	f
Negative	g	h	i

$$P(neg) = \frac{i}{(c+f+i)} \quad (1)$$

$$N(neg) = \frac{i}{(g+h+i)} \quad (2)$$

$$F1(neg) = \frac{2P(neg) \times R(neg)}{P(neg) + R(neg)} \quad (3)$$

$$A = \frac{a+e+i}{a+b+c+d+e+f+g+h+i} \quad (4)$$

2.3.3 N-gram for specific problem identification

To identify the precise nature of tourist dissatisfaction, word frequency analysis was conducted using the NLTK library (Bird et al., 2009). This study employed N-gram techniques (ranging from unigram to four-gram models) to determine the frequency of adjacent word sequences, specifically targeting phrases that reflect environmental and service-related impacts on marine and beach resources (Singgalen et al., 2024). The extracted N-grams were systematically categorized into three principal dimensions: Environmental (P_{Env}), Congestion/Access (P_{Acc}), and Service/Value (P_{Srv}), providing a structured dataset for geospatial planning and intervention.

2.4 Destination Sustainability Index (DSI) and hotspot analysis

2.4.1 Destination Sustainability Index (DSI)

The Destination Sustainability Index (DSI) is a composite index developed specifically for this study as a novel diagnostic tool, extending the core concept of the Tourist Satisfaction Index (TSI) (Gadekar, 2024). The DSI serves as a quantitative measure of geospatial sustainability risk based on the analysis of tourist dissatisfaction factors (Sari and Lestari, 2021) linked to physical and spatial management failures. It is constructed using the negative N-gram phrases extracted from reviews as input data (Ojo et al., 2021). The DSI calculation involves three steps:

(1) Creating the Binary Flag ($BF_{p,i}$): The process begins by calculating the Problem Density Ratio ($R_{p,i}$) for each dimension (p), which is the proportion of negative reviews mentioning problem p ($N_{p,i}$) relative to the total number of reviews (T_i).

$$R_{P,i} = \frac{N_{p,i}}{T_i} \quad (5)$$

(2) Defining the structural issue threshold: A problem density threshold of 5% was established based on the frequency distribution analysis of negative sentiments in large-scale social media datasets (Yuan, 2024). This operational decision is crucial to identify structural issues that require urgent policy intervention, ensuring that only recurring problems—rather than isolated incidents—are captured. The $R_{p,i}$ must exceed this level (corresponding to 5 out of every 100 reviews) for a problem dimension to be deemed structural. The

Binary Flag ($BF_{p,i}$) thus measures the appearance or non-appearance of the structural problem at destination i, assigned a value of 1 if $R_{p,i}$ >5% and 0 otherwise.

$$BF_{p,i} = \begin{cases} 1 \text{ if } R_{p,i} < 5\% \\ 0 \text{ if } R_{p,i} \leq 5\% \end{cases} \quad (6)$$

(3) DSI Calculation: The DSI_i value is calculated from the summation of the $BF_{p,i}$ across the three main problem dimensions (M=3: DSI_{Env}, DSI_{Acc}, and DSI_{Srv}). The index is structured as a normalization scale where a value close to 1.00 indicates high sustainability (low problem occurrence), and a value close to 0.00 indicates low sustainability (high problem occurrence) (adapted from Gadekar, 2024).

$$DSI_i = 1 - \frac{\sum_{p=1}^{M} BF_{p,i}}{M} \quad (7)$$

Where; DSI_i is the Destination Sustainability Index for destination i; $\sum_{p=1}^{M} BF_{p,1}$ is the summation of the Binary Flag (with values 0 or 1) across the three dimensions; and M represents the total number of problem dimensions, which is a constant value equal to 3.

2.4.2 Geospatial Hotspot mapping

To ensure the accurate and statistically significant identification of problem clusters, this research employs the hotspot analysis technique utilizing the G_i^* statistic (Getis and Ord, 1992). The DSI_i values derived from Equation (7) serve as the input variable (x_j), representing the attribute intensity analyzed at each location j. The spatial relationship is defined through a spatial weight matrix ($w_{i,j}$), and the G_i^* statistic assesses spatial clustering patterns as follow:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij}x_j - \bar{X}\sum_{j=1}^{n} w_{ij}}{S\sqrt{\frac{n\sum_{j=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij}\right)^2}{n-1}}} \quad (8)$$

Where; G_i^* is the G_i^* statistic value for destination i; x_j is the DSI_j value (input variable); $w_{i,j}$ is the spatial weight matrix value; n is the total number of destinations; $\bar{X}$ is the mean of all DSI values; and S is the standard deviation of all DSI values.

The G_i^* statistic generates a Z-Score and p-value to determine the significance of spatial associations. In this framework, a statistically significant negative Z-score (Z<-1.65 for 90% confidence level) indicates a spatial clustering of low DSI values (approaching 0.00), signifying a high sustainability risk (hotspot) characterized by high problem density. Conversely, a significant positive Z-score denotes a clustering of high DSI values (approaching 1.00), indicating high sustainability (coldspot). The resulting hotspot maps provide empirical geospatial evidence (Zhu and Newsam, 2024) for identifying critical zones within the Environmental, Congestion/Access, and Service/ Value dimensions. By differentiating significant hotspots by attraction type (BEACH and ISLAND), this approach maintains robust statistical power while offering precise evidence for targeted policy interventions and resource management (Singgalen et al., 2024).

2.5 Research framework and implementation

The research framework is designed as a sequential integration of deep learning and geospatial analysis to address tourism sustainability (Figure 2). The implementation commences with the collection and systematic pre-processing of geotagged reviews, followed by deep sentiment classification using the Pysentimiento model. Building upon the identified negative sentiments, N-gram analysis is employed for specific problem extraction, which serves as the fundamental input for calculating the Destination Sustainability Index (DSI) via the Binary Flag method. These indices are subsequently processed through geospatial Hotspot analysis using the Getis-Ord G_i^* statistic within a GIS environment (e.g., ArcGIS Pro or QGIS) to generate statistically significant problem maps. Finally, these spatial results are synthesized to formulate actionable sustainable tourism policies tailored for marine and beach attractions in Eastern Thailand. This integrated workflow ensures that complex textual data is effectively transformed into precise geospatial evidence for sustainability management.

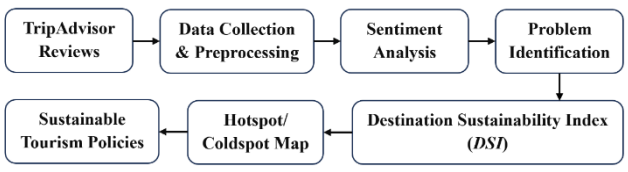


Figure 2. Integrated framework combining deep learning and geospatial analysis

3. RESULTS AND DISCUSSION

3.1 Sentiment analysis and model validation

3.1.1 Model performance and classification

The evaluation of the Pysentimiento model (BERT architecture) demonstrates robust classification performance, particularly within the negative class, which is central to this study. The model achieved a high precision value of 87%, confirming its reliability in accurately identifying reviews that contain specific dissatisfaction and management issues in the present study. While the overall classification accuracy was recorded at 0.76 due to the linguistic complexity of neutral sentiments, the high precision for the negative class (0.87) ensures that the subsequent geospatial analysis is based on highly reliable dissatisfaction data, minimizing false-positive errors in risk identification (Table 2).

From the total dataset of 11,047 reviews, the model classified 64.92% (7,172 reviews) as Positive, 16.52% (1,825 reviews) as Neutral, and 18.56% (2,050 reviews) as Negative. This distribution indicates that while the majority of tourists expressed satisfaction, nearly one-fifth of the digital feedback reflected structural or environmental concerns. Consequently, these 2,050 negative reviews were isolated to serve as the primary dataset for the subsequent N-gram feature extraction and Destination Sustainability Index (DSI) calculation. By focusing exclusively on this high-precision negative subset, the analysis ensures that the identified sustainability hotspots are derived from validated tourist dissatisfaction data, thereby enhancing the sensitivity of the DSI in detecting localized management failures.

Table 2. Performance metrics of deep learning sentiment model

Actual/prediction	Positive	Neutral	Negative	Recall	F1-score	Total
Positive	235	27	3	0.89	0.86	265
Neutral	29	33	8	0.47	0.42	70
Negative	15	26	75	0.66	0.74	116
Precision	0.84	0.38	0.87			
Accuracy					0.76	451

3.1.2 Spatial distribution of negative reviews

The visual concentration of negative feedback is illustrated in Figure 3, revealing a distinct clustering around high-density tourist hubs. Dissatisfaction is most prominent in the urbanized coastal zones near Pattaya City in Chonburi Province, as well as major marine destinations in Rayong and Trat Provinces. This initial spatial screening serves as the baseline for identifying potential risk locations. By providing a broad overview of dissatisfaction density, this distribution justifies the transition to the Destination Sustainability Index (DSI) calculation and subsequent hotspot analysis to confirm the statistical significance of these emerging patterns. Specifically, this visual density provides the necessary context for the Getis-Ord G_i^* analysis, ensuring that identified hotspots reflect true spatial dependencies rather than isolated incidents of negative feedback.

3.2 Specific problem identification and Destination Sustainability Index (DSI)

3.2.1 N-gram analysis and problem categorization

The N-gram analysis successfully isolated specific dissatisfaction clusters, confirming that while complaints are diverse, Environmental (P_{Env}) and Congestion/Access (P_{Acc}) issues are the most prominent concerns (Table 3). Environmental concerns are the most frequent, with 50% of the top 10 terms directly related to waste and hygiene (e.g., plastic, rubbish, dirty). This highlights waste management failures as a primary disruptor of the tourist experience. Similarly, Congestion/Access issues—consistently reflected through terms like "boat" and "crowded"—emerge as major dissatisfaction drivers. Although Service/Value (P_{Srv}) issues (e.g., "restaurant", "expensive") appear less frequently, they indicate underlying concerns regarding service quality and pricing. All extracted phrases were systematically categorized (Table 3) to serve as the functional foundation for calculating the Binary Flag ($BF_{p,i}$) and the subsequent DSI.

3.2.2 Destination Sustainability Index (DSI) overview

The Destination Sustainability Index (*DSI*) provides a quantitative synthesis of sustainability risks across 45 beaches and 12 islands (Table 4). Calculated via the Binary Flag ($BF_{p,i}$) method Equation (7), the results indicate that ISLAND attractions possess a

higher average sustainability level (0.47) compared to BEACH attractions (0.40). Despite this relative difference, both categories remain significantly below the ideal threshold, confirming that nearly all investigated sites face at least one structural problem dimension exceeding the 5% density threshold.

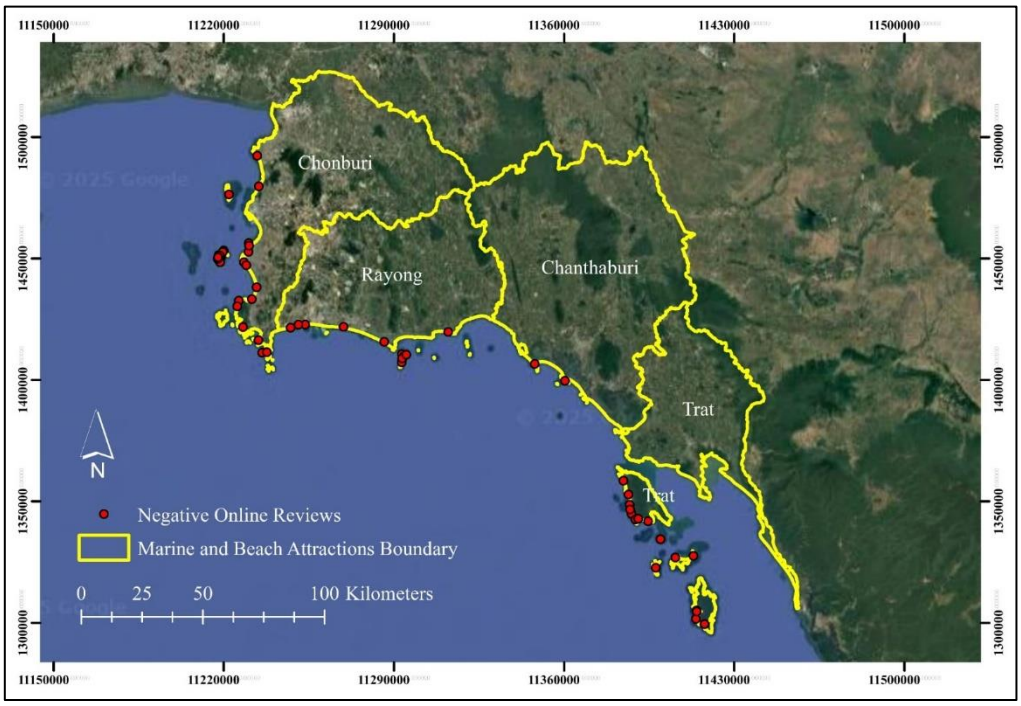


Figure 3. Spatial distribution of negative reviews

Table 3. Categorization of problem related N-grams and definition problem types (P)

Problem categories (P)	Description	High-frequency N-gram example
Environmental (P_{Env})	Physical quality and sustainability (waste, pollution, cleanliness)	plastic, rubbish, dirty, garbage, trash, bottle, filthy, litter, smell, polluted
Congestion/access (P_{Acc})	Space management, transport capacity, traffic	tourist, boat, crowded, road, noisy, street, busy, jetski, taxi, bike
Service/value (P_{Srv})	Service quality, price, feeling unsafe/unfair	restaurant, resort, hotel, vendor, shop, charge, expensive, staff, dog, police

Table 4. Average DSI values by problem dimension and attraction type

Type of attractions	$DSI_{Overall}$ (%)	DSI_{Env} (%)	DSI_{Acc} (%)	DSI_{Srv} (%)
BEACH	0.40	0.40	0.38	0.42
ISLAND	0.47	0.58	0.42	0.42

Dimension-specific analysis reveals critical structural profiles: Environmental quality (DSI_{Env}) serves as a comparative strength for the ISLAND group (0.58), suggesting that over half of these destinations remain free from structural environmental crises. Conversely, the BEACH group exhibits severe structural weakness in this dimension (0.40), reflecting widespread ecological degradation. The most acute risk identified is Congestion/Access (DSI_{Acc}) for the BEACH group (0.38), the lowest value in the dataset, which underscores the extreme vulnerability of easily accessible coastal mainland sites. Meanwhile, service/value (DSI_{Srv}) risks appear uniform across both groups (0.42), indicating a systemic challenge in service quality management that transcends attraction types.

3.3 Geospatial Hotspot mapping and policy implication

The application of Getis-Ord G_i^* statistics transforms numerical $DSI_{Overall}$ results into empirical geospatial evidence, pinpointing locations where sustainability risks are significantly clustered (Figure 4). The analysis confirms that these risks are not randomly distributed but exhibit a distinct spatial spectrum, ranging from high-sustainability clusters (coldspots) at 95%-99% confidence to emerging risk areas (hotspots) at 90% confidence (Table 5).

The strongest coldspot (99% confidence, $p<0.01$) was identified at Ao Phrao (Island, Trat), where a perfect $DSI_{Overall}$ of 1.00 was recorded. This cluster serves as a statistically significant "best practice management model," demonstrating effective problem control across all dimensions. Primary coldspots at 95% confidence were also observed across the beach group in Trat Province (average DSI 0.67), suggesting a regional trend of higher sustainability in the easternmost coastal zone. Conversely, a significant emerging hotspot (90% confidence) was detected near the Chonburi coastline, specifically encompassing Koh Si Chang and Koh Loi. These areas exhibit low DSI scores (0.00 to 0.67), indicating pervasive and dense sustainability challenges. From a policy perspective, this 90% confidence cluster requires immediate intervention and monitoring to prevent its escalation into a high-confidence sustainability crisis.

Notably, the majority of investigated destinations (e.g., Koh Kham, Chao Lao Beach, Wong Amat Beach) were categorized as not significant (Gi_Bin=0). Although some sites within this group reported low $DSI_{Overall}$ values (e.g., Koh Larn, Ao Wai), their dissatisfaction is not statistically clustered in a holistic sense. This suggests that while Koh Larn faces significant local challenges—particularly in service quality as detailed later—these issues do not yet constitute a regional-scale spatial contagion in the overall sustainability index. These "localized problems" suggest site-specific issues resolvable through internal management, rather than requiring broader regional-scale policy interventions.

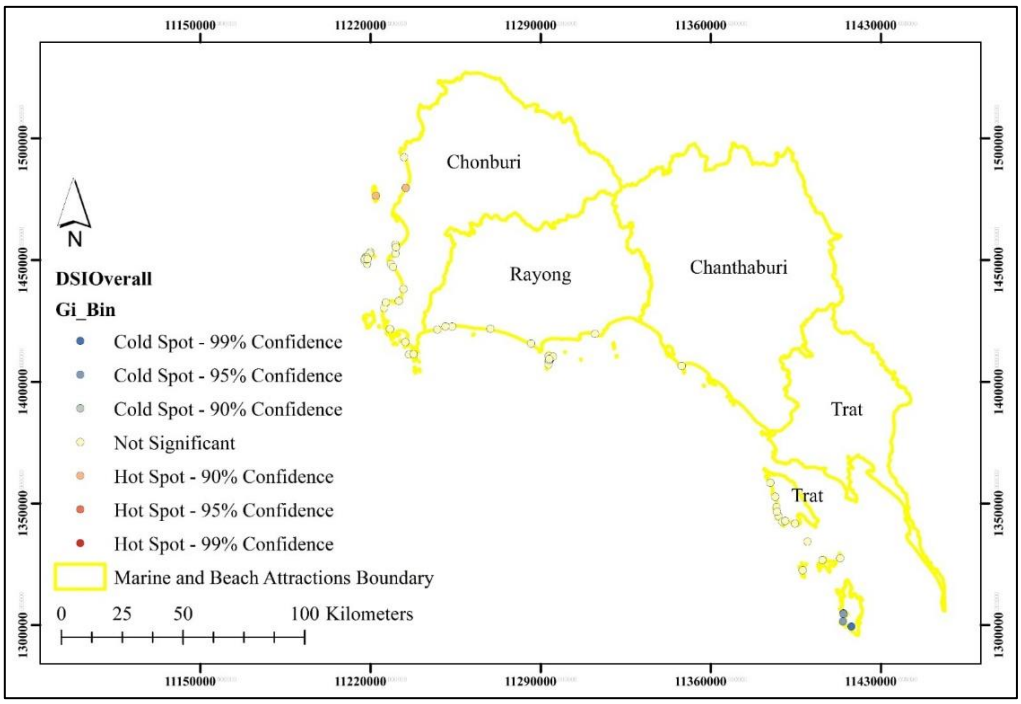


Figure 4. Geospatial hotspot analysis of overall sustainability ($DSI_{Overall}$)

3.4 Dimension-specific geospatial risk assessment

To provide granular insights for policy formulation, hotspot analysis was disaggregated into three core sustainability dimensions: Environment (P_{Env}), (P_{Acc}), and Service (P_{Srv}).

3.4.1. Environmental sustainability (DSI_{Env})

Results for P_{Env} reveal a unique spatial pattern where no statistically significant clusters were detected (Gi_Bin=0 for all locations). While sites like Pattaya and Koh Samed reported low local DSI_{Env} (0.00), these remained localized issues without

systematic spatial contagion. However, an "emerging risk" was identified in the Koh Samed cluster (Gi_Zscore ≈ -1.599, p ≈ 0.10), categorized as the highest watchlist area. This finding highlights the DSI's capability as an early warning tool, identifying forming clusters of degradation before they reach critical statistical significance (Figure 5, Table 6).

Table 5. Overall hotspot/coldspot summery ($DSI_{Overall}$)

Cluster type	Confidence (Gi_Bin)	Cluster area	Strategic policy implication
Strongest coldspot	99% (Gi_Bin=+3)	Ao Phrao Beach (Trat)	Management model: Replicate this "best practice" for high-value sustainability.
Primary coldspot	95% (Gi_Bin=+2)	Trat Beaches group (Avg. DSI: 0.67)	Standard support: Maintain high sustainability in coastal beach zones.
Emerging hotspot	90% (Gi_Bin=-1)	Koh Si Chang, Koh Loi (Chonburi)	Urgent watchlist: Immediate intervention required to prevent risk escalation.
Not significant	p>0.10 (0) (Gi_Bin=0)	Most destinations	Localized management: Address site-specific issues via internal measures.

Table 6. Summary of dimension-specific hotspot/coldspot analysis and management strategies

Problem dimension (P)	Cluster type	Confidence level	Primary locations	Strategic policy implication
Environmental (P_{Env})	Emerging risk	Approaching 90% (Z ≈ -1.60)	Koh Samed group (Rayong)	Highest watchlist: Implement early-warning waste management and proactive pollution control.
	Not significant	p>0.10	Most destinations	Maintain localized environmental standards; no regional contagion detected.
Congestion/access (P_{Acc})	Strongest coldspot	99% (Gi_Bin=+3)	Ao Phrao (Trat)	Management model: High-quality infrastructure; replicate this "low-density" model elsewhere.
	Primary hotspot	95% (Gi_Bin=-2)	Koh Si Chang, Koh Loi (Chonburi); White Sand Beach (Trat)	High crisis: Severe utility/transport deficits. Requires urgent structural investment and capacity limits.
	Significant coldspot	90% (Gi_Bin=+1)	Chao Lao Beach (Chanthaburi)	Sustainable access: Promote effective traffic and facility management as a secondary model.
Service/value (P_{Srv})	Primary hotspot	95% (Gi_Bin=-2)	Wonnapha and Wong Amat Beach (Chonburi)	Systemic failure: Critical service quality/safety risks. Enforce strict operator regulations and quality audits.
	Primary coldspot	95% (Gi_Bin=+2)	Ao Phrao (Trat)	Premium zone: Exceptional service standards. Designate as a benchmark for high-value tourism.
	Significant hotspot	90% (Gi_Bin=-1)	Koh Larn and Koh Si Chang clusters	Risk area: Rapid growth outpaces quality control. Urgent need for operator certification and service upgrades.

3.4.2 Congestion/access sustainability (DSI_{ACC})

Spatial clustering for P_{Acc} exhibits a sharp contrast between high-sustainability coldspots in Trat and high-risk hotspots in Chonburi (Figure 6, Table 6). The strongest coldspot (99% confidence) at Ao Phrao (Koh Kood) confirms that its infrastructure management serves as an ultimate best-practice model. Conversely, the upper Chonburi cluster (Koh Si Chang and Koh Loi) and the White Sand Beach (Trat) were identified as significant hotspots (95% confidence). These areas suffer from structural congestion where tourist volumes far exceed existing utility and transport capacities, requiring urgent infrastructure reinvestment.

3.4.3 Service/value sustainability (DSI_{Srv})

Analysis of P_{Srv} indicates systemic service-related failures clustered in high-density zones (Figure 7, Table 6). Significant hotspots (95% confidence) at Wonnapha and Wong Amat beaches signify recurring

issues in vendor quality and safety management. In contrast, the Koh Kood group maintained its status as a model zone (95% coldspot), emphasizing quality-over-quantity development. The identification of P_{Srv} hotspots in areas undergoing rapid growth, such as Koh Larn, suggests that service quality control often lags behind mass-transit expansion, necessitating stricter operator regulations.

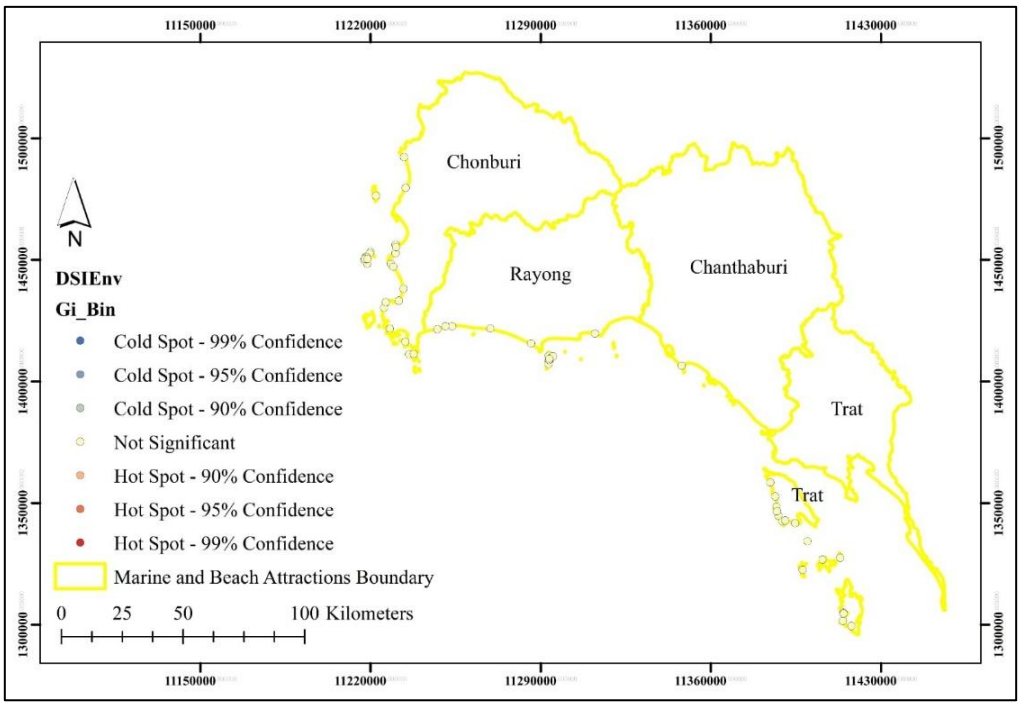


Figure 5. Geospatial hotspot analysis of Environmental sustainability (DSI_{Env})

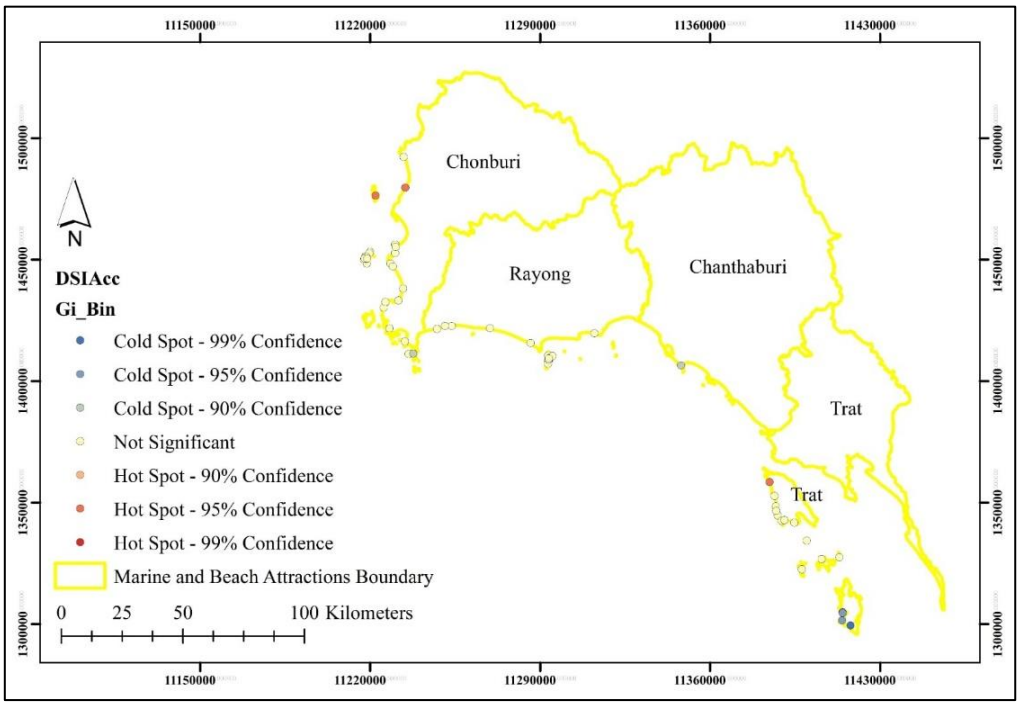


Figure 6. Geospatial hotspot analysis of Congestion/Access sustainability (DSI_{Acc})

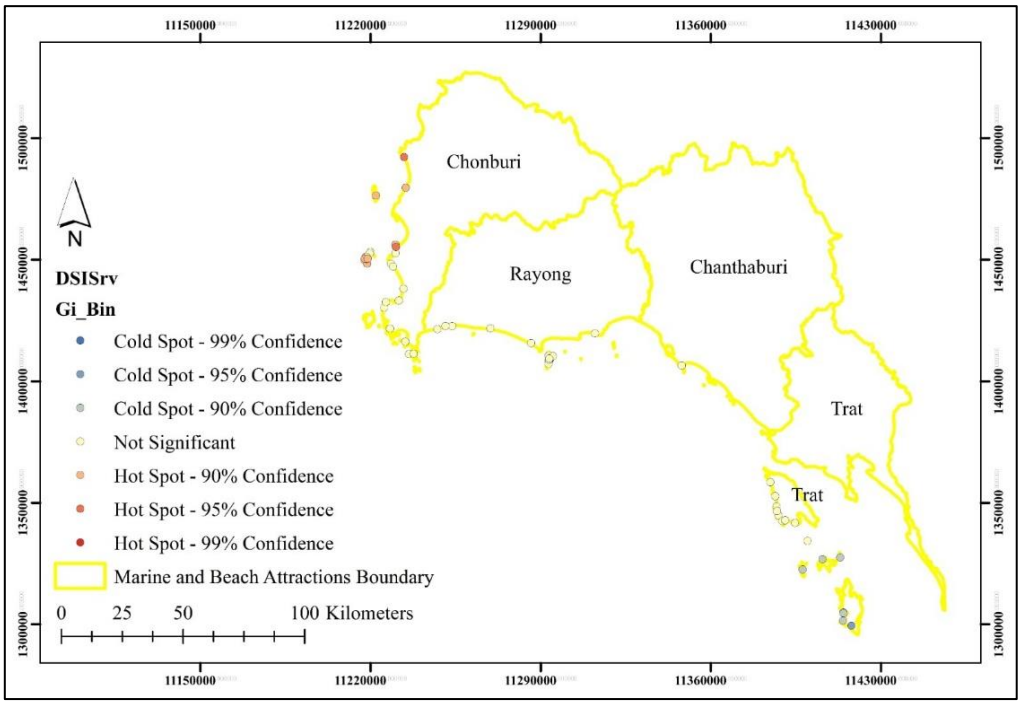


Figure 7. Geospatial hotspot analysis of service/value sustainability (DSI_{Srv})

3.5 Synthesis of geospatial risks and strategic policy implications

Based on the multi-dimensional hotspot analysis, the sustainability profile of Eastern Thailand's marine attractions can be synthesized into three strategic management domains:

(1) Critical intervention zones (priority hotspots): The Chonburi cluster (specifically Pattaya, Koh Si Chang, and Koh Loi) is identified as the highest priority area due to simultaneous Hotspot clustering in DSI_{Acc} and DSI_{Srv} dimensions. This overlap signifies a "complex structural crisis" where rapid, unmanaged growth has outpaced infrastructure capacity and service quality. While Koh Larn demonstrates high-density localized risks in the service dimension, the statistical G_i^* analysis prioritizes the Chonburi mainland and near-shore islands for regional-scale intervention. Urgent policy intervention is required, specifically focusing on strategic infrastructure investment—upgrading piers and waste systems while establishing rigid carrying capacity ceilings and tourism management fees to mitigate congestion.

(2) Model preservation zones (structural strengths): The Koh Kood cluster (Trat Province) consistently exhibits coldspot clustering, serving as the region's best practice management model. Policies for this zone must shift from development to preservation and scaling. By designating these areas as "premium sustainable zones," the government can provide incentives to operators while utilizing their low-density, high-quality management blueprint as a benchmark for other emerging destinations.

(3) Risk-based resource allocation: The Gi_Bin results provide an empirical foundation for optimized budget distribution. Locations categorized as significant hotspots (Gi_Bin≤-2) should be prioritized for structural budget allocation to rectify utility and transport deficits. Conversely, significant coldspots (Gi_Bin≥+2) should receive funding for sustainability maintenance and "quality-focused" marketing. For not significant areas with local issues, decentralized management is recommended to resolve site-specific problems without requiring regional-scale intervention.

4. CONCLUSION

This research successfully established a robust geospatial methodology for identifying "priority hotspots" of unsustainability in Eastern Thailand's marine destinations. By integrating a fine-tuned transformer-based BERT model (Pysentimiento) with the novel Destination Sustainability Index (DSI) and Getis-Ord G_i^* statistics, this study transformed 11,047 tourist reviews into empirical geospatial evidence. The model's high precision (87%) in classifying negative sentiments ensured that the 5% density threshold effectively isolated structural problems from localized

noise, providing a reliable foundation for regional policy formulation.

Key findings: (1) Dimensional vulnerability: N-gram analysis confirmed that environmental degradation (waste management) and congestion are the primary drivers of tourist dissatisfaction. The $DSI_{Overall}$ remains low across both attraction types (Island: 0.47, Beach: 0.40), with environmental quality (DSI_{Env}: 0.58) identified as the sole structural strength of the island group. (2) Spatial significance: The G_i^* analysis verified that sustainability risks are not randomly distributed but spatially clustered. The Chonburi cluster (specifically Koh Si Chang, and Koh Loi) emerged as the highest priority hotspot due to simultaneous failures in accessibility (DSI_{Acc}) and service quality (DSI_{Srv}). Conversely, the Koh Kood cluster (Trat) was consistently identified as a structural coldspot, representing a high-sustainability benchmark for the region.

ACKNOWLEDGEMENTS

The authors would like to thank the TripAdvisor website for the valuable information used in this research. Special thanks to the reviewers for their constructive comments and suggestions, which helped to improve the quality of this manuscript.

AUTHOR CONTRIBUTIONS

Methodology, software, formal analysis, data curation, validation, investigation, resources, visualization, and writing-original draft preparation, T.C.; writing-review and editing, N.P.; conceptualization-AI, A.S.; conceptualization-tourism, P.N. All authors have read and agreed to the published version of the manuscript

DECLARATION OF CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Abirami A. Marine pollution and waste management. Journal of Law and Legal Research Development 2024;1(2):20-4.

Alfred R, Chen Z, Eboy O, Luxuan Z, Renjie L. Analyzing trends in the spatial-temporal visitation patterns of mainland Chinese tourists in Sabah, Malaysia based on Weibo social big data. Heliyon 2023;11(15):e43981.

Baltranaite E, Inácio M, Pinto LV, Bogdziewicz K, Rocha J, Gomes E, et al. Tourism impacts on marine and coastal ecosystem services: A systematic review. Geography and Sustainability 2025;6(2):Article No. 100277.

Bird S, Klein E, Loper E. Natural Language Processing with Python. CA, USA: O'Reilly Media, Inc.; 2009.

Chakriswaran P, Vincent DR, Srinivasan K, Srinivasan K, Sharma V, Chang CY, et al. Emotion AI-driven sentiment analysis: A survey, future research directions, and open issues. Applied Sciences 2019;9(24):Article No. 5462.

Charfaoui K, Mussard S. Sentiment analysis for tourism insights: A machine learning approach. Stats 2024;7(4):1527-39.

Cui Y, He Q. Travel Behavior Analysis with Social Media Data and Smartphone GPS Data. NY, USA: University Transportation Center, University of Buffalo; 2019.

Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2019; Minneapolis, MN: USA; 2019.

Department of Tourism (DOT). Natural Tourist Attraction Quality Standard Assessment Manual. Office of Tourism Development [Internet]. 2018 [cited 2025 Nov 20]. Available from: https://www.dot.go.th/ebooks/ebooks-view/417.

Gadekar D. A study of tourist satisfaction index (TSI) of heritage tourist place of Ajanta Cave in Chatrapati Sambhaji Nagar District of Maharashtra State in India. Heritage: Journal of Multidisciplinary Studies in Archaeology 2024;11(2):1195-9.

Getis A, Ord K. The analysis of spatial association by use of distance statistics. Geographical Analysis 1992;24(3):189-206.

Jabalameli S, Xu Y, Shetty S. Spatial and sentiment analysis of public opinion toward COVID-19 pandemic using twitter data: At the early stage of vaccination. International Journal of Disaster Risk Reduction 2022;80:Article No. 103204.

Jeacle I, Carter C. In TripAdvisor we trust: Rankings, calculative regimes and abstract systems. Accounting Organizations and Society 2011;36(4):293-309.

Li X, Pan B, Law R, Huang X. Forecasting tourism demand with composite search index. Tourism Management 2017;59: 57-66.

Ministry of Tourism and Sports Thailand (MOTS). National Tourism Development Plan No. 3 (2023-2027) [Internet]. 2023 [cited 2025 Nov 20]. Available from: https://online.anyflip.com/zzfck/rptc/mobile/index.html.

Muthukadan B. Selenium Python Bindings [Internet]. 2024 [cited 2025 Oct 7]. Available from: https://selenium-python. readthedocs.io/.

Ojo O, Gelbukh A, Calvo H, Adebanji OO. Performance study of N-grams in the analysis of sentiments. Journal of the Nigerian Society of Physical Sciences 2021;3(4):477-83.

Office of Natural Resources and Environmental Policy and Planning (ONEP). Environmental Quality Situation Report 2024 [Internet]. 2024 [cited 2025 Oct 7]. Available from: https://www.onep.go.th/book/soe2567/.

Pérez JM, Rajngewerc M, Giudici JC, Furman DA, Luque F, Alemany LA, et al. Pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks [Internet]. 2023 [cited 2025 Oct 7]. Available from: https://doi.org/10.21203/rs.3.rs-3570648/v1.

Pleerux N, Boonpook W, Chaiyo A, Nardkulpat A. Trip advisor restaurant customer reviews analysis in Bangkok, Thailand. Proceedings of the International Conference on Cyber Management and Engineering (CyMaEn); 2023 Jan 26-27; Bangkok, Thailand; 2023.

Ramadhani A, Sutoyo E, Widartha VP. LSTM-based deep learning architecture of tourist review in Tripadvisor. Proceedings of the Sixth International Conference on Informatics and Computing (ICIC); 2021 Nov 3-4; Jakarta: Indonesia; 2021.

Sari PI, Lestari YD. Determinants of tourist satisfaction and dissatisfaction on tourism village. Journal Pendidikan Ekonomi Dan Bisnis 2021;9(1):9-24.

Singgalen YA, Wahyuningtyas SY, Widodo YE, Dasra MNA, Setiawan RW. Sentiment analysis for tourism destination reviews using cross-industry standard process for data-minning. Journal of Theoretical and Applied Information Technology 2024;102(17):6364-79.

Tourism Authority of Thailand (TAT). Survey Project for Analysis of International Tourist Behavior 2024 [Internet]. 2024 [cited 2025 Nov 20]. Available from: https://intelligencecenter.tat.or.th/articles/12311#tab-2.

Valdivia A, Luzón M, Herrera F. Sentiment analysis on TripAdvisor: Are there inconsistencies in user reviews? Proceedings of the International Conference on Hybrid Artificial Intelligence Systems (HAIS); June 21-23; La Rioja: Spain; 2017.

Xi Z, Min L, Yadong S. Night tourism satisfaction in the Qinghefang tourism and Leisure block based on an improved Kano Model. Journal of Resources and Ecology 2024; 15(1):55-65.

Yuan X. Evaluation of rural tourism development level using BERT-enhanced deep learning model and BP algorithm. Scientific Reports 2024;14:Article No. 25748.

Yusoh MP, Mapjabil J, Hanafi N, Idris MAM. Tourism carrying capacity and Social Carrying capacity: A literature review. SHS Web of Conferences 2021;124(4):Article No. 02004.

Zhu Y, Newsam S. Spatio-temporal sentiment hotspot detection using geotagged photos. Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems; 2016 Oct 31-Nov 3; Burlingame, CA: USA; 2016.