

Lexicon-Based Sentiment Analysis of Social Media Reviews for Floating Market Popularity in South Kalimantan

Evi Lestari Pratiwi¹, Ramadhani Noor Pratama², Inayatul Ulya Ahyati³, Paula Dewanti⁴
Study Program of Information System, Politeknik Negeri Banjarmasin, Banjarmasin, Indonesia^{1, 2, 3}
Department of Information Systems, Institut Teknologi dan Bisnis STIKOM Bali, Bali, Indonesia⁴

Abstract—Floating markets in South Kalimantan are culturally significant heritage destinations whose contemporary reputation is increasingly shaped by user-generated content on digital platforms. This study analyzes public perceptions of these markets by applying a lexicon-based sentiment analysis framework to 300 reviews collected from TripAdvisor and Google Maps between 2023 and 2024. The analytical workflow included text normalization, tokenization, stop-word removal, and stemming, with feature representation generated through term frequency-inverse document frequency (TF-IDF). Sentiment polarity was determined using bilingual lexicon-based scoring and categorized into positive, neutral, or negative sentiments. The results indicate that 45% of reviews expressed positive sentiment, highlighting cultural distinctiveness and riverfront experiences; 35% were neutral and provided descriptive logistical information; and 20% were negative, emphasizing waste issues, overcrowding, pricing inconsistencies, and perceived reductions in authenticity. TripAdvisor reviews exhibited greater emotional polarization than those on Google Maps. The findings demonstrate that lexicon-based sentiment analysis offers a transparent and effective approach for multilingual tourism contexts, providing insights into how digital narratives contribute to destination image formation. The study offers practical implications for improving environmental management, regulating visitor flows, and enhancing communication transparency within heritage tourism settings. It also contributes theoretically by underscoring the informational role of neutral reviews within electronic word-of-mouth dynamics. Future work may integrate machine learning-based sentiment classifiers or multimodal data to enhance analytical precision and extend the applicability of sentiment analysis in digital tourism research.

Keywords—Sentiment analysis; lexicon-based methods; floating markets; social media analytics; cultural tourism

I. INTRODUCTION

Tourism is one of the most rapidly expanding sectors in Indonesia, contributing approximately 4.3% to the national GDP in 2022 [1]. In 2023, the country welcomed more than 8.5 million international tourists, marking a strong post-pandemic recovery [2]. Within this national context, South Kalimantan continues to strengthen its tourism position, recording over 3.2 million domestic and international visitors in 2023 and generating substantial socio-economic benefits for local communities [3]. Among its key attractions are the floating markets, particularly the Lok Baintan and Kuin Floating Markets, which serve not only as centers of trade but also as

living cultural heritage that showcase the river-based traditions of the Banjar community. Government reports indicate that these markets account for nearly 25% of cultural tourism visits in the province [4].

The increasing influence of digital platforms such as TripAdvisor and Google Maps has reshaped how destinations are evaluated and promoted. User-generated content (UGC), especially online reviews, functions as digital word-of-mouth capable of shaping visitors' expectations, satisfaction levels, and overall destination image. Although several global studies have applied sentiment analysis to tourism contexts, research focusing on Indonesian heritage tourism, particularly floating markets, remains limited. Existing studies have not sufficiently explored multilingual reviews or applied computational approaches to understand emerging sustainability and service-related issues. This gap highlights the need for systematic analysis of tourists' online expressions to guide evidence-based destination management.

Although floating markets in South Kalimantan are widely recognized as cultural landmarks, they remain understudied from a computational perspective, particularly in relation to how visitors discuss issues of authenticity, service quality, and sustainability online. This gap underscores the need for a systematic understanding of how these destinations are portrayed in multilingual user reviews on major digital tourism platforms. Accordingly, this study seeks to explore how floating markets are perceived by examining the dominant positive, neutral, and negative sentiments expressed in visitor narratives. It aims to apply a lexicon-based sentiment analysis approach to multilingual reviews, identify key sentiment patterns and recurring themes, and translate these insights into practical recommendations for improving service delivery, sustainability efforts, and digital communication strategies.

The significance of this study is twofold. From an academic perspective, it enriches the growing body of work on sentiment analysis within Indonesia's cultural heritage tourism, demonstrating the value of lexicon-based methods for multilingual contexts and highlighting the often-overlooked role of neutral sentiment in shaping electronic word-of-mouth. From a practical standpoint, the findings provide guidance for policymakers, tourism authorities, and local communities seeking to enhance visitor experiences, manage crowding, address environmental concerns, and strengthen digital engagement.

II. LITERATURE REVIEW

A. Social Media and Tourism

Social media has reshaped the tourism landscape by enabling travelers to share photos, narratives, and evaluations in real-time (see Fig. 1). This user-generated content (UGC) fuels electronic word-of-mouth (e-WOM), which has become a key driver of trip planning, reputation building, and destination choice [5]. Xiang and Gretzel [6] emphasize that travelers increasingly rely on peer-generated information due to its perceived authenticity and credibility compared to official promotional materials. Positive online reviews can significantly enhance destination visibility and reputation, while negative reviews can reduce visitor interest or trust [7]. Empirical studies have shown that higher ratings and positive sentiment on platforms such as TripAdvisor correlate with increased bookings and stronger consumer confidence [8], positioning social media as both an opportunity and a reputational risk for tourism managers.



Fig. 1. Social media posting.

B. Sentiment Analysis in Tourism Research

Sentiment analysis, also referred to as opinion mining, is a branch of natural language processing (NLP) that seeks to determine the polarity (positive, neutral, negative) of textual data [9]. Liu [10] describes it as a computational technique for extracting opinions and emotions embedded in written language. Within tourism research, sentiment analysis enables large-scale examination of visitor experiences, allowing insights that go beyond traditional surveys or interviews [11] [12] [13] [14].

Recent studies have applied sentiment analysis to hotels, airlines, urban attractions, and destination branding, often using reviews from platforms like TripAdvisor and Yelp [15] [16] [17] [18]. While these methods have proven effective for monitoring satisfaction and guiding service improvements,

most studies rely on standard lexicon- or machine-learning approaches that prioritize polarity classification over contextual interpretation. Few studies extend sentiment findings to broader theoretical frameworks such as cultural sustainability or destination image formation, especially within developing-nation contexts (see Fig. 2).

A	B	C
229	Wisatawan biasanya mengambil foto di sekitar perahu.	netral
230	Tidak ada papan harga sehingga kadang wisatawan ditipu.	negatif
231	Harga makanan cukup murah dan pedagang sangat sopan.	positif
232	Cocok untuk berbelanja oleh-oleh atau melihat budaya lokal.	netral
233	Harga makanan cukup mahal dibandingkan pasar biasa.	negatif
234	Pasar terapung ini menjadi pengalaman yang berbeda dari tempat lain.	positif
235	Buka hanya di pagi hari, jadi sebaiknya datang lebih awal.	netral
236	Pasar sangat ramai, sulit bergerak, dan agak kotor di beberapa area.	negatif
237	Pasar terapung ini menjadi pengalaman yang berbeda dari tempat lain.	positif
238	Cocok untuk berbelanja oleh-oleh atau melihat budaya lokal.	netral
239	Pasar sangat ramai, sulit bergerak, dan agak kotor di beberapa area.	negatif
240	Sungguh destinasi wisata yang harus dikunjungi kalau ke Kalimantan.	positif
241	Akses ke pasar bisa menggunakan perahu motor atau klotok.	netral
242	Perahu sangat padat, terasa tidak nyaman.	negatif
243	Sangat bagus untuk wisata budaya, orang-orangnya ramah sekali.	positif
244	Akses ke pasar bisa menggunakan perahu motor atau klotok.	netral
245	Kebersihan kurang terjaga, perlu perbaikan.	negatif
246	Pasar terapung ini menjadi pengalaman yang berbeda dari tempat lain.	positif
247	Pasar terapung buka pagi hari, biasanya ramai menjelang subuh.	netral
248	Beberapa pedagang kurang ramah terhadap turis.	negatif
249	Makanan yang dijual enak dan khas daerah, wajib dicoba.	positif

Fig. 2. Sentiment analysis list.

C. Floating Markets as Tourism Icons

Floating markets are distinctive cultural and economic spaces found primarily in Southeast Asia (see Fig. 3). In Thailand, floating markets such as Damnoen Saduak have become iconic destinations attracting millions of visitors annually, although studies report concerns about overcrowding and commercialization [19]. Similarly, research in Vietnam identifies floating markets as culturally unique yet vulnerable to sustainability issues, including pollution and loss of authenticity [20] [21] [22].

In Indonesia, particularly South Kalimantan, floating markets such as Lok Baintan and Kuin serve dual purposes: they are active trade hubs for local communities and tourism attractions showcasing Banjar cultural heritage. Government tourism reports suggest that floating markets account for nearly 25% of cultural tourism visits in the province [3]. Despite their importance and visibility in tourism promotion, scholarly work on Indonesian floating markets has largely focused on cultural anthropology or local economic roles, with limited attention to how these sites are perceived online. Systematic digital-tourism analyses remain scarce.



Fig. 3. Floating markets as tourism icons.

D. Research Gap and Contribution

Although sentiment analysis is widely applied in tourism, two key gaps persist. First, cultural and heritage tourism,

particularly in developing regions, remains underrepresented, with few studies investigating how visitors interpret authenticity, sustainability, and service quality in heritage spaces such as floating markets [23] [24]. Second, existing studies often report polarity distributions without connecting results to deeper theoretical perspectives or policy relevance.

This study seeks to fill these gaps by offering three key contributions. First, on the methodological side, it demonstrates how a bilingual, lexicon-based sentiment analysis can be effectively applied to multilingual reviews in low-resource cultural tourism settings. Although lexicon-based approaches are well established, their tailored use for heritage tourism in Indonesia remains limited, making this application both relevant and responsive to the linguistic diversity and data constraints of the context. Second, at the theoretical level, the study connects sentiment patterns to broader discussions on destination image and cultural sustainability. It also advances e-WOM research by underscoring the often-underappreciated role of neutral sentiment, which can strengthen informational credibility and influence perceptions of authenticity. Third, the study offers practical value by outlining concrete strategies for tourism stakeholders to address waste management, manage crowd density, safeguard cultural authenticity, and enhance digital communication, key priorities for maintaining the long-term viability of heritage destinations in developing regions.

III. RESEARCH METHODS

This study advances current practices in digital tourism analytics by applying a bilingual lexicon-based sentiment analysis specifically tailored to multilingual heritage tourism reviews, a context that remains largely understudied in Indonesia. While lexicon-based techniques are not new, their customization for low-resource environments, where linguistic diversity is high and machine-learning training data are limited, represents a substantive methodological innovation. Unlike prior tourism studies that rely heavily on English-language or pre-trained classifiers, this research integrates Indonesian and English lexicons, adapts pre-processing rules to mixed-language narrative structures, and systematically examines sentiment polarity across platforms with differing user demographics. This tailored workflow, encompassing normalization, tokenization, stop-word refinement, bilingual stemming, and TF-IDF representation, demonstrates how lexicon-based models can maintain interpretability while achieving robust performance in multilingual, culturally embedded contexts. In doing so, the methodology provides a replicable and scalable framework for sentiment analysis in other data-sparse heritage tourism destinations.

This research employed a quantitative content analysis of user-generated content (UGC) sourced from TripAdvisor and Google Maps. A total of 300 reviews related to the two main floating markets in South Kalimantan—Lok Baintan and Kuin—were collected between January 2023 and December 2024. Reviews were purposively selected based on criteria such as language (Indonesian or English), clarity of visitor experiences, and the presence of at least one complete sentence to allow for sentiment analysis. Before classification, the dataset underwent a thorough preprocessing phase, which

included text normalization (case folding), removal of URLs, numbers, punctuation, and emojis, tokenization, stopword elimination in both languages, and stemming or lemmatization. Random checks ensured that these steps preserved the meaning of the original content.

Before conducting sentiment classification, the raw textual data underwent a comprehensive preprocessing pipeline using natural language processing (NLP) techniques. This process involved case folding to standardize text into lowercase, the removal of URLs, punctuation, numbers, emojis, and other noise elements, and tokenization to segment the reviews into lexical units. Stop-word removal was applied to eliminate high-frequency functional words in both Indonesian and English, while stemming or lemmatization was used to reduce words to their base forms. Manual spot-checking was conducted to ensure that semantic meaning was preserved following normalization.

For sentiment analysis, a lexicon-based approach was applied to categorize reviews into positive, neutral, or negative sentiments. Each token was assigned a polarity score (+1, -1, or 0), with negation handling implemented for words preceded by terms such as *tidak*, *bukan*, or *not*. A review's overall sentiment was determined by the aggregate score of its tokens. The findings were visualized using descriptive tables, pie and bar charts for sentiment distribution, as well as keyword frequency charts highlighting common terms in positive and negative reviews. To validate accuracy, 50 reviews were manually coded by two independent annotators, achieving an inter-rater reliability of 0.82 (Cohen's Kappa), which indicates strong agreement. Finally, a methodological flowchart was developed to illustrate the overall research process clearly.

The classified sentiment outcomes were subsequently examined through descriptive and visual analytics. Bar charts, pie charts, and word-frequency plots were produced to illustrate the distribution of sentiment categories and highlight dominant lexical patterns. Temporal comparisons between 2023 and 2024 were visualized using line charts to reveal changes in public perception, and platform-based differences were analyzed using comparative tables.

To ensure the reliability of the sentiment classification, a manual validation procedure was conducted. A subset of 50 reviews was independently annotated by two human coders with expertise in tourism communication and text analysis. The degree of agreement between the coders was assessed using Cohen's Kappa, which yielded a coefficient of 0.82, indicating strong inter-rater reliability. This validation step confirmed that the lexicon-based model performed consistently with human judgment, thereby supporting the credibility of the analytical results.

This study employed a structured methodological framework consisting of five primary stages, namely data collection, preprocessing, sentiment classification, visualization, and validation. The overall workflow followed in this research is illustrated in the methodological flowchart presented as Fig. 4. Each stage was designed to ensure analytical rigor, reproducibility, and alignment with established standards in computational tourism analytics.

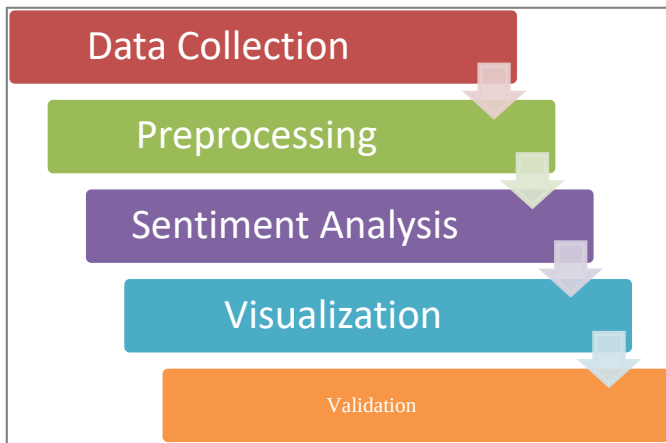


Fig. 4. Flowchart of research methods.

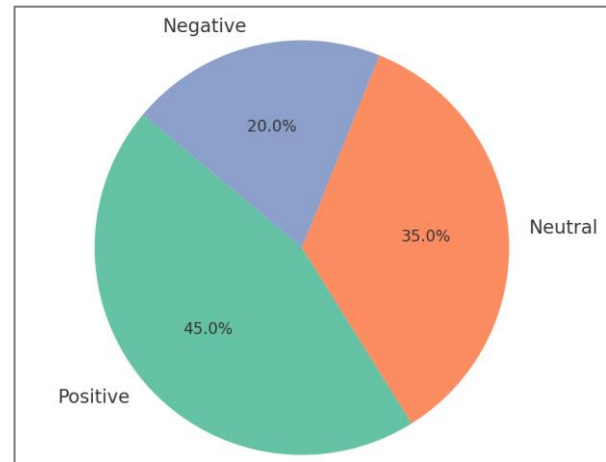


Fig. 5. Sentiment distribution (Pie chart).

IV. RESULTS AND DISCUSSION

A. Descriptive Statistics of the Dataset

The dataset comprised a total of 300 online reviews, collected from two primary platforms widely used in tourism research: Google Maps and TripAdvisor. Of these, 180 reviews (60%) originated from Google Maps, while 120 reviews (40%) came from TripAdvisor. In terms of language distribution, the majority of reviews, 210 reviews (70%), were written in Indonesian, reflecting local or domestic visitor experiences, while 90 reviews (30%) were in English, mostly representing international visitors.

The average length of reviews was 28.6 words per review, with English reviews being longer (average of 35 words) compared to Indonesian reviews (average of 25 words). This suggests that international tourists tend to write more detailed and elaborate reviews, possibly because they view their visit as a rare cultural experience worth documenting, whereas domestic tourists often provide shorter, more factual, and practical evaluations.

TABLE I. DATASET CHARACTERISTICS

Platform	Reviews	%	Avg. Length (words)
Google Maps	180	60%	27.1
TripAdvisor	120	40%	31.2
Total	300	100%	28.6

B. Sentiment Distribution

Sentiment classification divided the dataset into three categories: 135 reviews (45%) positive, 105 reviews (35%) neutral, and 60 reviews (20%) negative.

TABLE II. OVERALL SENTIMENT DISTRIBUTION

Sentiment	Frequency	Percentage
Positive	135	45%
Neutral	105	35%
Negative	60	20%

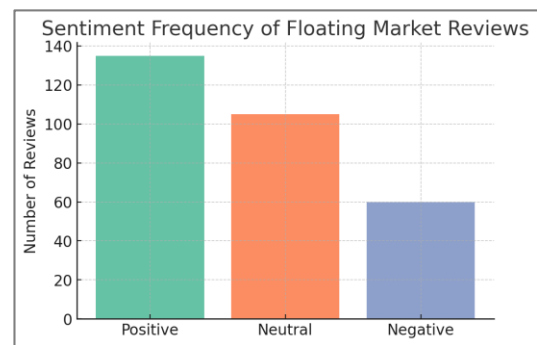


Fig. 6. Sentiment frequency (Bar chart).

Table I shows the dataset characteristics, while the sentiment distribution is presented in Table II. The results reveal that 135 reviews (45%) were positive, 105 reviews (35%) were neutral, and 60 reviews (20%) were negative. Fig. 5 and Fig. 6 illustrate the distribution through a pie chart and a bar chart. The dominance of positive reviews indicates a generally favorable perception of the floating markets, but the high percentage of neutral reviews suggests that many visitors provide descriptive rather than emotional feedback.

The results reveal that the floating markets are generally perceived positively by visitors, but a significant portion of reviews are neutral, indicating that many tourists use online platforms to provide objective descriptions rather than subjective evaluations. The presence of 20% negative reviews, however, highlights areas for improvement that may significantly influence future visitor perceptions.

C. Sentiment Analysis by Platform and Language

The sentiment analysis was disaggregated to compare results across the two platforms (Google Maps and TripAdvisor) and across the two primary languages used in the dataset (Indonesian and English). This breakdown reveals important differences in how visitors express their perceptions, shaped by both the nature of the platform and the cultural background of reviewers [25].

TABLE III. SENTIMENT ANALYSIS BY PLATFORM

Platform	Positive	Neutral	Negative	Total
Google Maps	75 (41.6%)	70 (38.8%)	35 (19.4%)	180
TripAdvisor	60 (50.0%)	35 (29.2%)	25 (20.8%)	120

TABLE IV. SENTIMENT ANALYSIS BY LANGUAGE

Platform	Positive	Neutral	Negative	Total
Google Maps	75 (41.6%)	70 (38.8%)	35 (19.4%)	180
TripAdvisor	60 (50.0%)	35 (29.2%)	25 (20.8%)	120

The platform-based analysis shows that TripAdvisor reviews were more positive compared to Google Maps, likely because TripAdvisor attracts more international tourists who emphasize cultural and scenic uniqueness. In contrast, Google Maps reviews leaned more neutral, suggesting that local visitors are more likely to report factual observations (such as directions, operating hours, or crowd levels) instead of strong opinions (see Table III).

The language-based analysis indicates that English reviews are more polarized, with higher proportions of both positive and negative comments. This can be explained by cultural

differences in review-writing behavior: international visitors tend to express enthusiasm when satisfied but are also more explicit when disappointed. On the other hand, Indonesian reviews are predominantly neutral, showing a preference for descriptive narratives rather than extreme sentiment (see Table IV).

D. Keyword and Topic Analysis

Keyword analysis identified recurring terms in each sentiment category (see Fig. 7):

1) Positive reviews emphasized words such as beautiful, unique, friendly, clean, and morning. Many of these highlighted the cultural authenticity of the floating market, the friendliness of local traders, and the picturesque atmosphere during sunrise.

2) Neutral reviews commonly contained words such as location, boat, access, and time. These were descriptive and provided practical information, such as transportation modes, market schedules, and general features of the market.

3) Negative reviews frequently included words like crowded, dirty, expensive, and traffic. These highlight operational problems such as overcrowding during peak hours, waste management challenges, and inflated tourist prices.

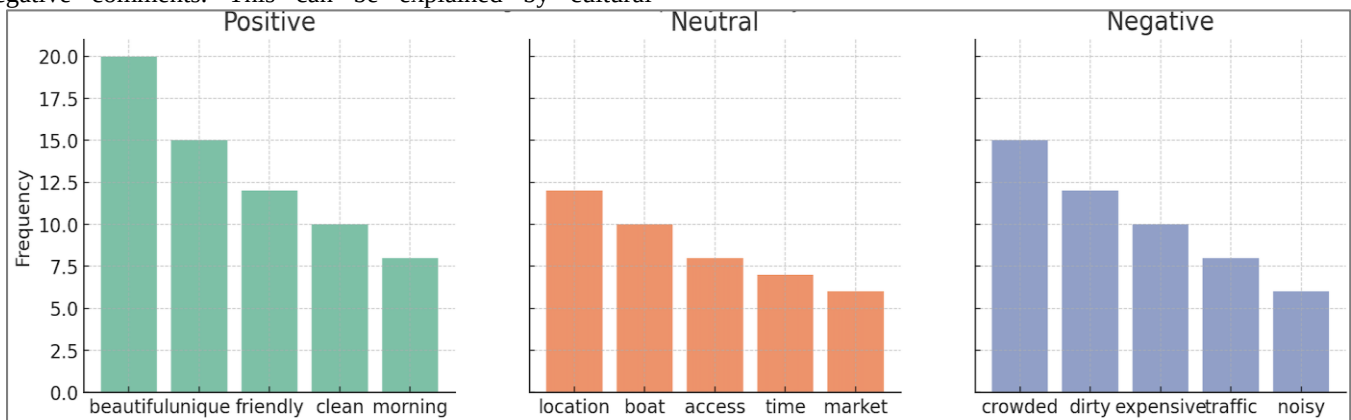


Fig. 7. Keyword and topic analysis (Bar chart).

E. Temporal Trends (2023 versus 2024)

A year-to-year comparison provides additional insights into how perceptions of the floating markets evolved between 2023 and 2024. Table V presents the sentiment distribution for both years, followed by Fig. 8 which visualizes the temporal trend. The data shows a clear shift in visitor perceptions, with positive sentiment decreasing and negative sentiment increasing over time.

TABLE V. SENTIMENT DISTRIBUTION BY YEAR

Year	Positive	Neutral	Negative	Total
2023	90 (50%)	63 (35%)	27 (15%)	180
2024	45 (37.5%)	42 (35%)	33 (27.5%)	120

In 2023, visitor perceptions of the floating markets were largely favorable, with about half of the reviews highlighting their cultural authenticity. Tourists frequently emphasized the distinctive experience of trading from boats, the picturesque

sunrise views, and the warmth of local vendors. Negative feedback was minimal (15%) and generally limited to minor concerns such as crowded mornings or a lack of adequate facilities.

By contrast, in 2024 the sentiment shifted noticeably. Positive reviews fell to 37.5%, while negative reviews rose to 27.5%, reflecting growing dissatisfaction with waste management, congestion on the river, and perceptions of inflated prices for tourists. Although neutral reviews remained stable at around 35%—providing practical guidance on transport, schedules, and directions—the overall pattern suggests a mixed outcome. Social media-driven popularity increased tourist arrivals, yet at the same time exposed weaknesses in infrastructure and service delivery. This rising share of negative sentiment underscores the urgency for better crowd control, enhanced environmental management, and transparent pricing policies to prevent the floating markets from losing their authenticity and being perceived as over-commercialized.

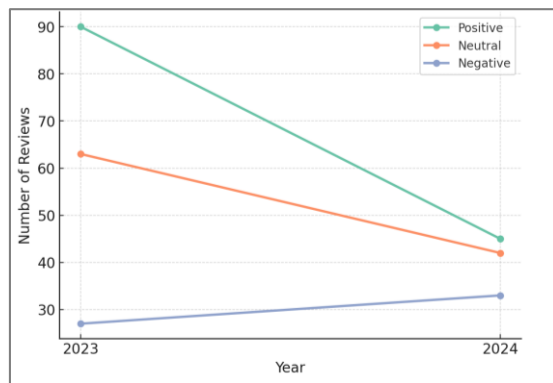


Fig. 8. Temporal trends of sentiment (2023 vs. 2024).

F. Interpretation of Neutral Reviews

Neutral reviews, which accounted for 35% of the dataset, primarily contained factual and descriptive information rather than emotional expressions. Visitors often mentioned elements such as accessibility, opening hours, transportation by boat, or general atmosphere without attaching strong judgments. For instance, one reviewer wrote, *“The market is open early in the morning, and it is better to come before 7 a.m. to avoid crowds”*, while another stated, *“Boats are available from the pier, and the trip to the market takes about 15 minutes”*. These examples show how neutral reviews serve as practical guides that help potential visitors plan their trips.

The relatively high proportion of neutral reviews also contributes to the credibility and authenticity of online platforms. While positive and negative reviews reflect emotions, neutral reviews provide balance by offering straightforward accounts of the market. Their stability across both years indicates that they function as a baseline narrative, consistently delivering reliable context regardless of changes in sentiment trends. From a managerial perspective, these reviews highlight recurring concerns about logistics and access, signaling that clearer communication in promotional materials could enhance the visitor experience.

G. Negative Sentiment and Service Gaps

Negative reviews, although representing only 20% of the dataset, carry significant weight in shaping the overall perception of the floating markets. These reviews frequently pointed out recurring problems such as waste accumulation on the river, overcrowding of boats during peak hours, and perceptions of unfair pricing targeted at tourists. For example, one visitor noted, *“The river is beautiful, but there was a lot of floating trash which spoiled the experience”*, while another commented, *“Too many boats at the same time made it uncomfortable to enjoy the scenery”*. Such feedback highlights that operational issues, rather than cultural or aesthetic shortcomings, are the main drivers of dissatisfaction.

From a managerial perspective, these negative reviews reveal critical service gaps that demand immediate attention. Waste management must be prioritized to preserve both environmental quality and visitor impressions, while crowd control strategies—such as staggered visiting hours or regulated boat traffic—could reduce congestion. Addressing concerns about pricing transparency is also essential, as the

perception of unfair treatment can damage trust and deter repeat visits. If left unresolved, these issues risk eroding the long-term sustainability of the floating markets, potentially overshadowing their cultural uniqueness and reducing their appeal to both domestic and international audiences.

H. Social Media as a Double-Edged Sword

Social media operates as a dual-impact mechanism in shaping the digital reputation of South Kalimantan’s floating markets. On the positive side, large volumes of favorable and neutral user-generated content function as organic promotion. Reviews and images emphasizing cultural authenticity, vibrant morning activity, and trader hospitality amplify destination visibility across platforms such as TripAdvisor, Google Maps, Instagram, and TikTok. These narratives significantly influence travel decision-making processes and act as low-cost digital marketing channels [26].

Negative reviews spread rapidly and can harm reputation, if ignored. Complaints about crowding, sanitation, and pricing easily circulate online, discouraging visits. Continuous monitoring through automated sentiment analysis and real-time dashboards enables managers to balance positive promotion with early warnings, ensuring adaptive strategies and sustainable competitiveness in digital tourism.

I. Managerial Implications

The findings of this study provide actionable insights for integrating sentiment analysis into digital tourism management. First, the predominance of positive reviews suggests opportunities for leveraging data-driven promotional strategies. Destination managers can employ automated content curation and recommender systems to amplify user-generated narratives across social media platforms, thereby enhancing visibility and engagement.

Second, the recurrent negative sentiments regarding waste, overcrowding, and pricing indicate specific service quality gaps that can be addressed using data-informed decision support systems. For instance, predictive analytics could assist in regulating visitor flow, while digital monitoring tools may support transparency in pricing and operational efficiency.

Third, local community engagement can be strengthened through capacity-building programs supported by digital platforms. Training traders in online promotion and eco-friendly practices, coupled with access to analytics dashboards, can preserve authenticity while improving competitiveness.

Finally, the adoption of smart tourism technologies, such as real-time sentiment monitoring and automated alerting systems, would allow destination managers to detect emerging issues proactively. This approach demonstrates how computational sentiment analysis can function as a decision-support mechanism, positioning heritage destinations like South Kalimantan’s floating markets as benchmarks for sustainable, technology-enabled tourism management.

J. Theoretical Implications

This study contributes theoretically by extending sentiment analysis to cultural heritage tourism, specifically floating markets in South Kalimantan. Using lexicon-based methods [27], it shows how computational models capture cultural,

economic, and environmental dynamics. The significant share of neutral reviews reveals that e-WOM is both emotional and informational, with neutral content serving as a stabilizing factor for credibility.

Additionally, differences identified across platforms and languages underline the necessity of adopting cross-platform and multilingual approaches. The contrast between English-dominated TripAdvisor and Indonesian-dominated Google Maps indicates that cultural communication styles and platform features shape sentiment expression. Temporal analysis between 2023 and 2024 further shows that sentiment is dynamic, shifting alongside infrastructure development, sustainability challenges, and digital exposure. These insights refine sentiment analysis theory by emphasizing contextual, linguistic, and temporal factors in tourism-focused applications.

This study has several limitations that should be acknowledged. The dataset was restricted to text-based reviews from TripAdvisor and Google Maps, which may not represent perceptions expressed on multimedia platforms such as Instagram or TikTok. The lexicon-based method, although transparent and suitable for multilingual data, may not fully capture nuanced expressions such as sarcasm or implicit sentiment. The analysis also relied solely on textual content, excluding visual cues that influence tourist evaluations. Furthermore, the two-year timeframe may not reflect longer-term sentiment shifts. Future studies should incorporate cross-platform, multimodal, and machine learning-based approaches for broader insight.

Although this study offers substantial insights into public perceptions of South Kalimantan's floating markets, several limitations should be noted. The dataset, drawn solely from TripAdvisor and Google Maps, excludes richer multimedia formats from platforms such as Instagram or TikTok that may reveal additional experiential dimensions. The lexicon-based sentiment approach, despite its transparency and suitability for multilingual analysis, may not fully capture subtle linguistic nuances such as irony, implicit sentiment, or culturally specific expressions, particularly within mixed Indonesian-English reviews. The two-year scope (2023 to 2024) may also limit the ability to observe longer-term sentiment fluctuations. Nevertheless, the study contributes significantly by extending sentiment analysis into the underexplored field of cultural heritage tourism and by introducing a cross-linguistic analytical framework that enhances understanding of how multilingual digital narratives shape cultural sustainability. These insights help clarify how online discourse both reflects and influences the preservation and perceived authenticity of heritage destinations.

V. CONCLUSION AND FUTURE WORK

This study employed a bilingual lexicon-based sentiment analysis to examine 300 online reviews from Google Maps and TripAdvisor, generating insights into the public's digital perception of South Kalimantan's floating markets. The findings reveal a sentiment distribution of 45% positive, 35% neutral, and 20% negative. Positive content emphasizes cultural richness and river-based experiences, while negative reviews highlight persistent challenges related to sanitation, congestion,

pricing inconsistencies, and concerns over declining authenticity. Cross-platform differences show that TripAdvisor reviews, largely posted by international visitors, exhibit greater emotional polarization, whereas Google Maps reviews tend to be more neutral and informational. Temporal patterns also point to rising negative sentiment in 2024, signaling emerging sustainability pressures that require immediate intervention.

Beyond describing sentiment patterns, this study offers meaningful insights for tourism strategy, policy development, and cultural preservation. The results highlight how UGC shapes the destination image of floating markets, influencing visitor expectations and perceptions of cultural sustainability. From a managerial standpoint, addressing waste, crowding, and pricing transparency is essential for safeguarding authenticity and improving overall visitor satisfaction. These findings support data-driven decision-making by enabling tourism authorities to monitor real-time visitor sentiment, guide targeted environmental management, regulate visitor capacity, and refine digital communication strategies to promote more accurate and appealing representations of the markets.

From a theoretical perspective, the study expands the application of sentiment analysis by demonstrating how neutral sentiment, often ignored in tourism analytics, plays a key role in conveying factual, credibility-enhancing information within e-WOM. This insight contributes to broader discussions on destination image formation and digital heritage interpretation, particularly in developing-nation contexts where cultural tourism holds strategic socioeconomic value.

Looking ahead, future research can advance this work by integrating machine-learning or hybrid sentiment models, incorporating multimodal data (e.g., images, videos), and examining cross-cultural differences in perception. Such developments can deepen analytical precision and further strengthen the role of sentiment analysis in heritage tourism governance and digital tourism planning.

Future work may extend this study by integrating larger and more diverse datasets from additional social media platforms such as Instagram or TikTok. Employing advanced machine learning methods, including transformer-based deep learning models, could improve classification accuracy and capture nuanced emotional expressions beyond lexicon constraints. Longitudinal analyses with extended timeframes would enable more robust trend detection, while cross-country comparative studies could provide generalizable insights into sentiment dynamics across cultural heritage tourism domains

REFERENCES

- [1] UNWTO. International Tourism Highlights 2023. Madrid: World Tourism Organization; 2023.
- [2] Badan Pusat Statistik. Statistik Wisatawan Mancanegara 2023. Jakarta: BPS; 2024.
- [3] Dinas Pariwisata Provinsi Kalimantan Selatan. Laporan Kunjungan Wisatawan 2023. Banjarmasin: Dispar Kalsel; 2024.
- [4] Kementerian Pariwisata dan Ekonomi Kreatif RI. Statistik Pariwisata Indonesia 2023. Jakarta: Kemenparekraf; 2024.
- [5] Litvin SW, Goldsmith RE, Pan B. Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*. 2008;29(3):458–468.

- [6] Xiang Z, Gretzel U. Role of social media in online travel information search. *Tourism Management*. 2010;31(2):179–188. doi:10.1016/j.tourman.2009.02.016.
- [7] Sparks BA, Browning V. The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*. 2011;32(6):1310–1323. doi:10.1016/j.tourman.2010.12.011.
- [8] Filieri R, McLeay F. E-WOM and accommodation: Factors influencing travelers' adoption of information. *Journal of Travel Research*. 2014;53(1):44–57. doi:10.1177/0047287513481274.
- [9] Pang B, Lee L. Opinion mining and sentiment analysis. *Found Trends Inf Retr*. 2008;2(1–2):1–135. doi:10.1561/1500000011.
- [10] Liu B. *Sentiment Analysis and Opinion Mining*. San Rafael (CA): Morgan & Claypool; 2012.
- [11] Marine-Roig E, Clavé SA. Tourism analytics with massive user-generated content: A case study of Barcelona. *J Destin Mark Manag*. 2020;18:100469. doi:10.1016/j.jdmm.2020.100469.
- [12] Al-Smadi M, Talafha B, Al-Ayyoub M. Sentiment analysis using term-based methods for Amazon product reviews. *Int J Adv Comput Sci Appl*. 2022;13(1):685–691.
- [13] Arifiyanto D, Ferdiana R. Sentiment analysis on customer satisfaction of digital banking in Indonesia. *Int J Adv Comput Sci Appl*. 2022;13(5):466–473.
- [14] Alsheikh S, Elhag H. Detection of sentiment polarity of multi-language text from social media. *Int J Adv Comput Sci Appl*. 2018;9(6):199–203.
- [15] Erdoğan D, Aksu U, Esen FE. Developing a deep learning-based sentiment analysis system of hotel customer reviews for sustainable tourism. *Sustainability*. 2025;17:5756. doi:10.3390/su17135756.
- [16] Charfaoui K, Laachfoubi A, Hain M. Sentiment analysis for tourism insights: A machine learning approach to TripAdvisor reviews. *Informatics*. 2024;7(4):90. doi:10.3390/informatics7040090.
- [17] Apriliani D, Abidin T, Sutanta E, Hamzah A, Somantri O. Sentiment analysis for assessment of hotel services review using decision tree feature selection. *Int J Adv Comput Sci Appl*. 2020;11(4):240–245. doi:10.14569/IJACSA.2020.0110432.
- [18] Budianto A. Natural language processing for the analysis sentiment using an LSTM model. *Int J Adv Comput Sci Appl*. 2022;13(3):777–785. doi:10.14569/IJACSA.2022.01303100.
- [19] Cohen E, Avieli N. Food in tourism: Attraction and impediment. *Ann Tourism Res*. 2004;31(4):755–778. doi:10.1016/j.annals.2004.02.003.
- [20] Phong N, Nguyen H. Sentiment analysis of online reviews in tourism: Case of Vietnam. *Asian J Tourism Res*. 2020;5(1):12–25.
- [21] Mendes GR, Oliveira T, Lima L, dos Santos DE, Kontogianni M, Arvanitis S, et al. Cultural heritage tourism and sustainability: A bibliometric analysis. *Sustainability*. 2024;16:6424. doi:10.3390/su16156424.
- [22] Tazi A, Bouarfa M, Oussous A, Kenitar S, Serhani MA. Exploring destination's negative e-reputation using aspect-based sentiment analysis: Case of Marrakech. *Tour Manag Perspect*. 2021;40:100892.
- [23] Kumar C, Singh M. Sentiment analysis for tourism insights: A machine learning approach. *Stats*. 2024;7:1527–1539.
- [24] Nadzirah I, Karim I, Fathi M, Mustafa R, Su'ud MM. Exploring tourist experience through online reviews using zero-shot aspect-based sentiment analysis. *Information*. 2024;15:499. doi:10.3390/info15080499.
- [25] Zovko I, Šimunović F, Matić H, Bago B. A sentiment analysis model based on user experiences of Dubrovnik on TripAdvisor. *Appl Sci*. 2024;14:8304. doi:10.3390/app14188304.
- [26] Martins G. The online destination image portrayed by user-generated content and its impact on tourist engagement. *Tourism & Management Studies*. 2024;20(3):XX–XX.
- [27] Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. *Comput Linguist*. 2011;37(2):267–307. doi:10.1162/COLI_a_00049.