

Explainable Aspect-Based Sentiment Analysis for Marine Tourism Using Large-Scale Language Models

Aisah Rini Susanti^{1*}, Hadiyanto², Ridwan Sanjaya³

Doctoral Program in Information Systems, School of Postgraduate Studies, Diponegoro University, Semarang, Indonesia¹

Department of Computer Science-Faculty of Computer Science, Djuanda University, Bogor, Indonesia¹

Department of Chemical Engineering-Faculty of Engineering, Diponegoro University, Semarang, Indonesia²

Department of Information Systems-Faculty of Computer Science, Soegijapranata Catholic University, Semarang, Indonesia³

Abstract—Marine tourism is a strategic sector heavily influenced by tourists' perceptions and experiences, which are captured in user-generated content across various digital platforms. However, the multifaceted, informal, and ambiguous nature of reviews poses challenges for sentiment analysis based on conventional methods. This study proposes an aspect-based sentiment analysis (ABSA) framework using a Large Language Model (LLM) integrated with Explainable Artificial Intelligence (XAI) to analyze cross-platform marine tourism reviews. The proposed approach includes text reflection preprocessing, aspect extraction, sentiment classification based on zero-shot and few-shot learning schemes, and XAI integration to improve the interpretability of the results through the contribution of keywords and evidence phrases. Model evaluation is conducted by comparing the model output against reference annotations on a subset of review data using precision, recall, and F1-score metrics at the aspect level. Experimental results show that the zero-shot scheme produces moderate performance with an accuracy and macro-average F1 score of 0.33, while the application of some-shot learning significantly improves performance with an accuracy of 0.87 and a macro-average F1 score of 0.87. The integration of XAI shows that the resulting sentiment prediction model is supported by relevant linguistic evidence, thereby increasing transparency and trust in the analysis results. These findings confirm that the proposed ABSA-based LLM framework is effective, interpretative, and has the potential to be used as a decision-support tool in the management and development of data-driven marine tourism destinations.

Keywords—LLM; aspect-based sentiment analysis; XAI; marine tourism; UGC

I. INTRODUCTION

Marine tourism is a tourism subsector that has experienced significant growth in recent decades, along with increasing tourist interest in coastal destinations, islands, and marine ecosystem-based activities. The characteristics of marine tourism, which are highly dependent on environmental conditions, service quality, and safety factors, make the sector sensitive to tourist perceptions and experiences. These perceptions play a crucial role in shaping destination image, satisfaction levels, and repeat visit intentions, which ultimately determine competitiveness and tourism [1].

Digital transformation has changed the way travelers express their experiences and opinions. Digital platforms and social media allow travelers to share reviews openly through user-generated content (UGC) on platforms such as Google Maps,

TripAdvisor, Instagram, TikTok, and travel forums. UGC is now a key source of information in the travel decision-making process, while also representing authentic experiences of travelers in real time [2], [3], [4]. Therefore, analyzing UGC is becoming an increasingly important approach for destination managers and tourism stakeholders to understand tourist perceptions and continuously rank destination reputations [5].

However, analyzing marine tourism reviews faces complex challenges. Traveler reviews are generally unstructured, using informal language, slang, emojis, and mixed languages, and often contain more than one opinion in a single sentence. Travelers may express positive sentiments about one aspect, such as the beauty of a beach, but simultaneously convey negative sentiments about another aspect, such as accessibility or facilities. This linguistic complexity and multifaceted nature make it difficult for conventional sentiment analysis methods to accurately capture tourists' meanings and emotions [6].

Sentiment analysis has been widely applied in tourism research to extract tourists' opinions from digital reviews. Common approaches are dictionary-based (lexicon-based) [7], [8] or classical machine learning algorithms, such as Naïve Bayes [9], [10], [11] and Support Vector Machine [12], [13]. Although relatively easy to implement, these methods have limitations in understanding context, irony, and complex emotional nuances. The development of deep learning and transformer models, such as Convolutional Neural Network, Recurrent Neural Network, and BERT, has improved the acquisition of sentiment analysis [14], [15], [16]. However, these models generally require large labeled datasets and expensive training processes, making them less flexible for specific and dynamic domains such as marine tourism.

To limit sentiment analysis globally, Aspect-Based Sentiment Analysis (ABSA) was developed to identify specific aspects discussed in the review and determine the polarity of sentiment on each aspect [17], [18], [19], [20]. In the context of marine tourism, ABSA is very relevant because it allows for tourist guidance on key aspects, such as beach cleanliness, environmental conditions, safety, facilities, and accessibility. However, most ABSA research in tourism still relies on pre-defined dictionary aspects or supervised learning approaches, making it less adaptive to the ever-evolving variations in languages, digital platforms, and destination contexts.

The development of Large Language Models (LLMs) opens new opportunities in aspect-based sentiment analysis. LLMs,

*Corresponding author.

trained on large, multilingual text corpora, have superior capabilities in understanding context, capturing implicit meaning, and performing zero-shot and some-shot learning without the need for large labeled datasets [21], [22]. In tourism sentiment analysis, LLMs have proven to be more effective in handling informal language, complex emotions, and reviews containing multiple aspects within a single text [23]. However, LLMs also present new challenges, particularly related to their black-box nature, making the analysis results difficult to explain and interpret for non-technical users. The lack of transparency and interpretability are crucial issues in the application of LLM-based sentiment analysis, especially when the analysis results are used as a basis for managerial and public policy decisions. In the context of marine tourism, stakeholders require not only sentiment classification results, but also explanations regarding which aspects trigger certain sentiments and the underlying linguistic evidence. Therefore, the integration of Explainable Artificial Intelligence (XAI) becomes very important to increase trust, accountability, and usability [24], [25], [26].

Based on these gaps, this study proposes an Explainable Aspect-Based Sentiment Analysis approach for marine tourism using Large Language Models. The proposed approach integrates the contextual capabilities of LLMs with XAI mechanisms to produce aspect-based sentiment analysis that is not only accurate but also explainable. Through this approach, the system is able to identify key aspects of marine tourism, determine the sentiment polarity for each aspect, and provide explanations based on text evidence regarding the linguistic factors that influence the analysis results.

The main contribution of this research lies in the development of an aspect-based sentiment analysis approach based on a Large Language Model specifically designed to handle informal, multi-faceted, and highly contextual reviews of marine tourism characteristics. This approach is enriched with the integration of Explainable Artificial Intelligence principles to increase the transparency and interpretability of sentiment analysis results, so that the rationale behind the prediction model can be clearly understood. Furthermore, this research provides an applicable analytical foundation and can be utilized as a data-driven decision support tool for destination managers and marine tourism stakeholders in converging tourist perceptions and developing more effective destination management strategies.

II. METHOD

This study proposes an Explainable Aspect-Based Sentiment Analysis (ABSA) framework based on a Large Language Model (LLM) to analyze tourist perceptions of marine tourism destinations. The research methodology includes collecting marine tourism review data from various digital platforms, preprocessing review texts, identifying and extracting tourism aspects, LLM-based sentiment classification, integrating XAI mechanisms for model interpretability, and evaluating performance. The zero-shot approach performs sentiment classification without labeled examples, whereas the few-shot approach incorporates two labeled examples to guide the LLM in identifying tourism-related aspects and their corresponding sentiment polarities. The flow of the research methodology stages is shown in Fig. 1.



Fig. 1. Flow of stages of research methodology.

A. Data Collection

The data used in this study consists of user-generated content (UGC) related to marine tourism, collected from various digital platforms such as Google Maps Reviews, TripAdvisor, Instagram, YouTube, and other blog review platforms. Data were collected using web scraping techniques and official application programming interfaces (APIs), while adhering to privacy policies and ethical data usage. The reviews collected reflect tourists' opinions on various aspects of marine tourism destinations, such as the quality of the marine environment, tourist facilities, services, accessibility, cleanliness, and the overall experience. The data are multilingual and informal, reflecting the characteristics of user communication in digital media.

TABLE I. EXAMPLE SUMMARY OF MULTI-PLATFORM MARINE TOURISM REVIEW DATASET

Platform	Data source	Content Type	Number of Samples	Information
Google Maps	Google Maps Review	Text Review	3,200	Tourist reviews of beaches and marine tourism attractions
TripAdvisor	TripAdvisor Reviews	Text Review	2,150	Visitors' opinions about marine tourism destinations
Instagram	Hashtag Scraper	Description and Comments	1,800	User descriptions and comments using the marine tourism hashtag
YouTube	YouTube Comment Scraper	Comment	1,250	Viewer comments on marine tourism videos

Table I shows a summary of the marine tourism review dataset from various social media platforms. The study reports a total of 8,400 user-generated records obtained from four digital

platforms: Google Maps (3,200 records), TripAdvisor (2,150 records), Instagram (1,800 records), and YouTube (1,250 records). The collected data represent different forms of tourism-related user-generated content, including destination reviews and online comments. The implementation supports CSV-based data ingestion and the identification of relevant text columns. Each record is processed using a configurable preprocessing pipeline, and the application records the number of input records and the number retained after relevance filtering. The initial collection size is distinguished from the final analytical dataset size. Language distributions, duplicate counts, and record exclusions should be reported using the actual preprocessing logs and dataset records to ensure reproducibility.

B. Text Preprocessing

Text preprocessing is the initial stage aimed at improving the quality of data reviews before they are analyzed using a Large Language Model (LLM). Given the informal, multilingual nature of user-generated content (UGC) in marine tourism, which often contains non-textual symbols, the preprocessing process is designed in a streamlined manner to avoid losing semantic context that is crucial for model understanding.

This decision was made to preserve the sentence structure and semantic context necessary for the LLM to understand implicit meaning and relationships between words. Stopword removal was limited, only for words that did not contribute semantically to sentiment analysis. The final preprocessing stage was formatting the text into a prompt-ready input form that conforms to zero-shot and some-shot learning schemes. Each review was structured in a consistent input format, enabling the LLM to perform aspect identification and sentiment classification stably and replicably. Table II shows examples of text preprocessing output and a prompt for creation.

The preprocessing pipeline comprises configurable text-cleaning and normalization operations. First, malformed text is repaired using the *ftfy* library. HTML tags and URLs are removed when the corresponding options are enabled. Additional operations include optional emoji removal, case folding, and the normalization of Indonesian informal expressions using a predefined slang dictionary.

Language identification is performed using lexical heuristics to distinguish Indonesian, English, and unknown-language records. A supplementary heuristic identifies potentially code-mixed Indonesian–English expressions. Limited stopword removal is then applied using predefined language-specific stopword lists.

A marine-tourism relevance filter is used to identify records containing domain-related keywords associated with beaches, oceans, waves, islands, coral reefs, and marine activities. The application stores cleaned text, detected language, text length, and relevance indicators.

The preprocessing configuration must be recorded consistently for each experiment because text normalization and filtering may influence the information available for subsequent aspect and sentiment classification.

C. Aspect Extraction

The aspect extraction process was carried out to identify the key dimensions of the marine tourism experience discussed in each user review. This stage began with the use of review text that had undergone a preprocessing process to clean and normalize it without losing semantic context. Next, a collection of previously established marine tourism aspects was used as a reference framework to ensure the extracted aspects remained relevant and consistent.

Each review was then formulated into a structured prompt and analyzed using a Large Language Model (LLM), which is capable of understanding both explicit and implicit meaning in the text. In addition to identifying emerging aspects, the model also extracted supporting phrases (evidence phrases) that represent the basis for determining each aspect, thereby increasing transparency and explainability. Results. The output from this stage is presented in a structured format to support the aspect-based sentiment classification process and explainability analysis in the next stage. Fig. 2 presents a sample output from the aspect extraction results.

D. Sentiment Analysis Using LLM

Sentiment analysis using a Large Language Model (LLM) was conducted to classify sentiment polarity for each previously identified aspect of marine tourism. At this stage, preprocessed and aspect-extracted review texts were analyzed through a structured prompt modeled to determine positive, negative, or neutral sentiment at the aspect level. The aspect-based approach allows a single review to generate multiple different sentiment labels, corresponding to the multidimensional characteristics of the tourism experience. LLM was utilized for its ability to understand semantic context and implicit expressions commonly found in *user-generated content*. For each aspect, the model also extracted *evidence phrases* as the basis for determining sentiment, thereby increasing the transparency and explainability of the analysis results. Inference was performed using *zero-shot* and *some-shot learning schemes* to assess the consistency of sentiment classification, with the output presented in a structured format to support analysis and interpretation in subsequent stages.

E. Integrating Explainable AI

The integration of Explainable Artificial Intelligence (XAI) was carried out to improve the transparency and interpretability of sentiment analysis results based on Large Language Models (LLMs). In this study, the XAI approach is implemented through an evidence-based explanation mechanism, where each aspect and polarity of the generated sentiment is accompanied by evidence phrases taken directly from the review text.

F. Integrating Explainable AI

The integration of Explainable Artificial Intelligence (XAI) was carried out to improve the transparency and interpretability of sentiment analysis results based on Large Language Models (LLMs). In this study, the XAI approach is implemented through an evidence-based explanation mechanism, where each aspect and polarity of the generated sentiment is accompanied by evidence phrases taken directly from the review text. This approach allows for explicit tracing of the relationship between the input text and the model's decisions, thereby reducing the

black-box nature of LLMs. By presenting the analysis results in a structured format, the XAI integration also supports qualitative analysis and validation of the results, as well as increasing user

and stakeholder confidence in the proposed analytical system in the context of marine tourism.

Aspect Extraction

Runs: cleaning, case folding, slang normalization, limited stopwords, language detection, relevance filter, then produces prompt-ready outputs and a heuristic (offline) aspect extraction preview.

Text column
text
Choose the column that contains review/comment/caption text.

Options (toggle)

☒ remove_url

☒ remove_html

☒ remove_emoji

☒ case_folding

☒ slang_norm

☒ code_mixed

☒ limited_stopwords

☒ relevance_filter

☒ prompt_ready

☒ aspect_extraction

☒ heuristic_aspects

☒ absa_prompt

New toggles: aspect_extraction (enable aspect stage), heuristic_aspects (preview aspects without API), absa_prompt (export ABSA JSONL for LLM inference).

Run Batch

Download CSV

Export Prompts JSONL

Export ABSA JSONL

View Meta

Processed Preview (first 10 rows)

Copy buttons help you quickly test prompts in your LLM. Aspect extraction preview is heuristic (offline) for interpretability.

text_raw	text_clean_sw	language	is_relevant	aspects_json (preview)	prompt_aspect_extraction
Visited for 3 nights and enjoyed breakfast, lunch and dinner here each day. A great variety to choose from - seafood, Indonesian and western choices. The seafood bbq buffet on the Saturday night was the highlight. Staff very friendly and attentive. The food was made to order and very fresh and tasty. Made the trip that much better.	visited for 3 nights enjoyed breakfast, lunch dinner here each day. a great variety choose from - seafood, Indonesian and western choices. seafood bbq buffet on saturday night was highlight. staff very friendly attentive. food was made order very fresh tasty. made trip much better.	en	True	Show (2 aspects) [[{"aspect": "environment/sea quality", "evidence": "a great variety choose from - seafood, Indonesian western choices"}, {"aspect": "service", "evidence": "staff very friendly attentive"}]]	Copy Show You are an expert in aspect extraction for marine tourism reviews. Extract only the aspects from this list: environment/sea quality, cleanliness facilities, service, accessibility/parking, safety, price/value, crowding. For each extracted aspect, provide an evidence phrase from the text. Return a JSON array with fields: aspect, evidence. Text: visited for 3 nights enjoyed breakfast, lunch dinner here each day. a variety choose from - seafood, Indonesian western choices. seafood bbq buffet saturday night was highlight. staff very friendly attentive. food was made to order and very fresh tasty. made trip much better.
CBC is a great place to relax and chill away from the hustle and bustle of Jakarta. For surfers, it's paradise with an awesome surf break right off the shore. For non-surfers the food is delicious, the views are soothing and the cool breeze in the restaurant makes it a great place to relax, read a book, work or play games with friends. We really enjoyed playing with our son in the pool surrounded by greenery walking 10 minutes over the beach nearby look at the seashells and play in the tidal pools.	cbc a great place relax chill away from hustle bustle jakarta. for surfers, it's paradise with an awesome surf break right off shore. for non-surfers food delicious, views are soothing cool breeze restaurant makes a great place relax, read a book, work or play games with friends. we really enjoyed playing with our son pool surrounded by greenery walking 10 minutes over beach nearby look at seashells play tidal pools.	en	True	Show (1 aspects) [[{"aspect": "environment/sea quality", "evidence": "we really enjoyed playing with our son pool surrounded by greenery walking 10 minutes over beach nearby look at seashells play tidal pools"}]]	Copy Show You are an expert in aspect extraction for marine tourism reviews. Extract only the aspects from this list: environment/sea quality, cleanliness facilities, service, accessibility/parking, safety, price/value, crowding. For each extracted aspect, provide an evidence phrase from the text. Return a JSON array with fields: aspect, evidence. Text: cbc a great place relax chill away from hustle bustle jakarta. for surfers it's paradise with an awesome surf break right off shore. for non-surfers food delicious, views are soothing cool breeze restaurant makes a great place relax, read a book, work or play games with friends. we really enjoyed playing with our son pool surrounded by greenery walking 10 minutes over beach nearby look at seashells play tidal pools.

Fig. 2. Aspect extraction interface with example output and prompt generation for LLM-based ABSA.

G. Model Evaluation

An evaluation was conducted to assess the performance of the proposed LLM-based sentiment analysis framework in identifying aspects and classifying sentiments in marine tourism reviews.

Given the aspect-based and generative nature of the approach, the evaluation focused not only on sentiment classification accuracy but also on the quality of aspect extraction and explainability. The evaluation process was conducted by comparing the model output against manually compiled reference annotations on a subset of the review data. Each aspect-sentiment pair generated by the model was assessed

for aspect suitability, sentiment polarity, and the relevance of the provided evidence phrases. The evaluation was conducted using two inference schemes, namely zero-shot and some-shot learning, to measure the effect of the inclusion of examples on the stability and precision of the results. For quantitative aspects, the evaluation metrics used included precision, recall, and F1-score at the aspect level, which provide a balanced picture between criteria and completeness of predictions. In addition, a qualitative analysis was conducted by reviewing the evidence phrases generated by the model to assess the level of explainability and consistency of explanations, especially in cases of ambiguous reviews or reviews containing more than one aspect with different sentiment polarities.

TABLE II. EXAMPLE OF TEXT PREPROCESSING OUTPUT AND PROMPT GENERATION

Platform	Text (Raw)	Text Cleanup	Relevant	Command Types
Google Maps	The beach is very clean, and the water is crystal clear, but parking is limited during peak hours.	The beach is very clean, and the water is crystal clear, but parking is limited during peak hours.	CORRECT	Zero shots
TripAdvisor	The water is incredibly clear, and the waves are calm. The staff is friendly, but the entrance fee is quite expensive.	The water is incredibly clear, and the waves are calm; the staff is friendly, but the entrance fee is a bit expensive.	CORRECT	Few shots
Instagram	Amazing sunset view with crystal-clear water and a relaxing atmosphere #beach #snorkeling	Amazing sunset view with crystal clear water and relaxing atmosphere #beach #snorkeling	CORRECT	Zero shots
YouTube	The sea view is beautiful, but the road access to the location still needs improvement.	The sea view is beautiful, but the road access to the location still needs improvement.	CORRECT	Few shots

III. RESULTS AND DISCUSSION

A. Sentiment Analysis Results Using LLM

The evaluation was conducted on a manually annotated subset; therefore, future studies should consider larger and more diverse annotated datasets to further assess the generalizability of the proposed approach. The results of sentiment analysis using a Large Language Model (LLM) demonstrate that the proposed approach is able to consistently identify user sentiment tendencies in marine tourism reviews sourced from various social media platforms. LLM not only classifies sentiment polarity (positive, neutral, and negative) but also captures the nuances of the language context, which is informal, descriptive, and often contains subjective expressions typical of user-generated content. Overall, the analysis results show a dominance of neutral and positive sentiments, indicating that most users convey marine tourism experiences in an informative or descriptive manner, accompanied by positive evaluations of the quality of the environment, scenery, and recreational experiences offered.

The LLM-based approach also demonstrates stability in cross-platform data processing, despite significant variations in language characteristics, text length, and review delivery style. This confirms that LLMs are able to adapt to data heterogeneity without requiring rigid linguistic rules, making them suitable for use in multi-platform marine tourism sentiment analysis. Fig. 2 shows an example of the results of a cross-platform sentiment analysis for tax tourism using LLM.

Fig. 2 shows the distribution of sentiment analysis results across various social media platforms: Google Maps, Instagram, TripAdvisor, and YouTube. Neutral sentiment dominates across all platforms, particularly on Google Maps and YouTube, with a significantly higher number of neutral reviews than positive and negative ones. This pattern suggests that users on these platforms tend to provide factual and descriptive reviews, such as information about location, access, or the general condition of the destination, without extreme emotional expressions.

On Google Maps, while neutral sentiment dominates, positive sentiment is also relatively significant. This indicates that the platform is frequently used to share firsthand, evaluative

experiences, particularly regarding environmental quality and the natural beauty of marine tourism.

Negative sentiment on Google Maps is relatively low, indicating that user complaints or dissatisfaction only arise in certain aspects and do not dominate the overall perception.

Instagram shows fewer reviews than other platforms, with a relatively balanced sentiment distribution, but still dominated by neutral sentiment. This may be attributed to Instagram's visual nature, where descriptive text tends to be brief and complementary, limiting explicit sentiment.

TripAdvisor displays a significantly higher proportion of positive sentiment than negative sentiment. This finding is consistent with TripAdvisor's role as a travel review platform, where users generally write more reflectively and evaluatively about their experiences. Meanwhile, YouTube is dominated by neutral sentiment with a large number of reviews, reflecting the large number of comments in the form of short responses, questions, or general discussions about marine tourism video content.

B. Explainable AI (XAI) Integration Results

The integration of Explainable Artificial Intelligence (XAI) into a Large Language Model (LLM)-based sentiment analysis framework aims to increase the transparency and explainability of the model's decision-making process. Unlike conventional sentiment classification approaches, which tend to be black-box, the XAI approach in this study allows each sentiment result to be traced back to relevant pieces of review text as evidence. The XAI component provides evidence phrases and textual explanations to support the interpretation of aspect-level sentiment predictions. Although these explanations improve transparency, further quantitative evaluation is required to assess their faithfulness.

In its implementation, XAI interacts through an evidence-based explanation mechanism, where the LLM not only generates sentiment labels for each aspect but also extracts key phrases or sentences from original reviews that explicitly represent user assessments. Each aspect-sentiment pair is complemented by textual evidence derived directly from user-generated content, making the interpretation of the results more contextual and amenable to manual processing.

Explainable AI: Key Word Contributions Across All Sentiments

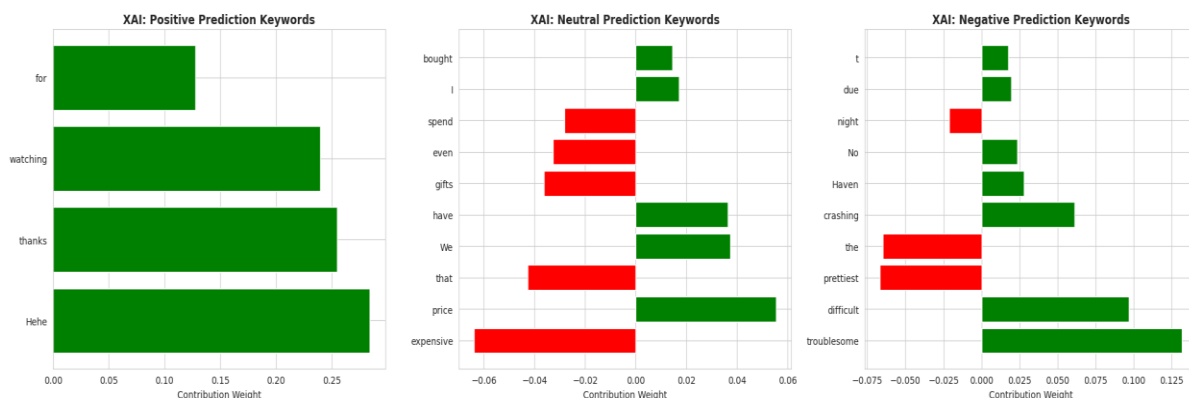


Fig. 3. Explainable AI (XAI) keyword contribution analysis for sentiment prediction.

Fig. 3 shows the results of the Explainable Artificial Intelligence (XAI) integration, which visualizes the contribution of keywords to the sentiment predictions generated by the model. Word contributions are categorized into three categories: positive, neutral, and negative sentiment, providing a more transparent picture. Regarding linguistic factors influencing decision models.

In the positive sentiment category, words like "Hehe," "thanks," and "watching" had high contribution weights, reflecting expressions of satisfaction and appreciation commonly found in marine tourism reviews. This indicates that the model is capable of recognizing informal expressions as indicators of positive sentiment, even though these words are not always directly related to the technical aspects of the service.

The neutral sentiment category displays a relatively balanced contribution of positive and negative words. Narrative words like "have" and "we" contribute little to neutral sentiment, while price-related terms like "expensive" and "price" tend to push sentiment toward the evaluative. These findings indicate that discussions of cost are a frequent trigger for sentiment changes, although not always explicitly negative.

For negative sentiment, words like "troublesome," "difficult," and "crashing" showed strong contributions, representing the user's experience of discomfort or difficulty. Conversely, general or positive words, such as "the" and "prettiest," tended to reduce the intensity of negative sentiment when they appeared in specific contexts.

C. Model Evaluation Results

After conducting sentiment analysis based on the Explainable Artificial Intelligence (XAI) integration aspects and mechanisms, the next step is to disseminate the quality and reflection of the model's results. This evaluation aims to assess the extent to which the sentiment analysis output consistently represents user perceptions, is contextually relevant, and can reliably support further analysis. Table III presents the model evaluation results for the Few-Shot Classification Report.

TABLE III. FEW-SHOT SENTIMENT CLASSIFICATION REPORT

Class	Accuracy	Remember	F1 Score	Support
Negative	0.85	0.96	0.90	23
Neutral	0.93	0.74	0.83	35
Positive	0.85	0.93	0.89	42
Accuracy	-	-	0.87	100
Macro Average	0.87	0.88	0.87	100
Weighted Average	0.88	0.87	0.87	100

Table III presents sentiment classification using the few-shot approach on marine tourism reviews using precision, recall, and F1-score metrics. The evaluation results show that the positive class has the highest precision value (0.40), indicating that the model is relatively more accurate in identifying positive sentiment. Conversely, the negative class has a lower precision value (0.24), reflecting the challenge in recognizing expressions of dissatisfaction that are often implicitly conveyed in user-generated content.

The overall accuracy value and macro-average F1-score of 0.33 indicate that the model's performance is still at a moderate level, which is a common characteristic of zero-shot approaches without a specific training domain. To gain a deeper understanding of the classification error patterns, the numerical results are further visualized as a confusion matrix in Fig. 4.

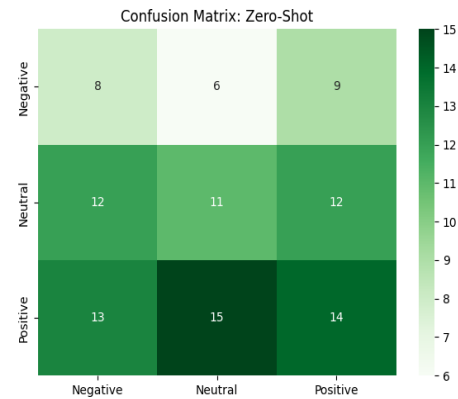


Fig. 4. Confusion matrix of sentiment classification without prior training (zero-shot).

Next, to assess the effect of including labeled examples in the inference process, an evaluation was conducted using a some-shot learning scheme. This approach aims to improve the model's understanding of the context and boundaries between sentiment classes by providing a number of representative examples within the prompt. Fig. 4 presents the results of the sentiment classification performance evaluation using the some-shot scheme in the form of a Zero-Shot Classification Report.

The evaluation results of some-shot learning show significant performance improvements compared to the zero-shot scheme. The accuracy value reached 0.87, with a macro-average F1-score of 0.87, indicating that the inclusion of labeled examples in the prompts helps the model produce more stable and consistent predictions across sentiment classes.

At the class level, negative sentiment demonstrated very strong performance with a recall of 0.96 and an F1-score of 0.90, indicating the model's high ability to comprehensively identify negative reviews. The positive class also displayed balanced performance, with a precision of 0.85, a recall of 0.93, and an F1-score of 0.89, reflecting the accuracy and completeness of predictions for positive sentiment. Meanwhile, the neutral class had the highest precision (0.93) but a lower recall (0.74), indicating that the model tends to be conservative in predicting neutral sentiment and more precise in targeting that class.

Overall, the improvement in precision, recall, and F1-score values of the some-shot scheme confirms that the inclusion of contextual examples helps the LLM model better understand the boundaries between sentiment classes, especially in multi-aspect and ambiguous marine tourism reviews. This finding strengthens the argument that few-shot learning provides substantial benefits in improving the performance of aspect-based sentiment classification without requiring full model retraining [27], [28]. Fig. 5 shows the confusion matrix of the Few-Short Classification Report.

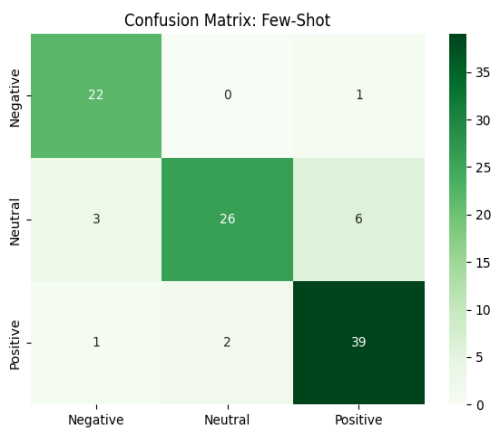


Fig. 5. Few-shot sentiment classification confusion matrix.

Overall, the research results show that the proposed Large Language Model (LLM)-based marine tourism sentiment analysis framework is capable of providing a comprehensive understanding of user perceptions across various social media platforms. The aspect-based sentiment analysis (ABSA) approach enables the identification of key aspects that shape the tourist experience, while the integration of Explainable Artificial Intelligence (XAI) enhances the transparency and explainability of the analysis results. This combination makes the model not only effective in generating predictive sentiment but also interpretively accountable.

The quantitative evaluation results show a clear performance difference between the zero-shot and some-shot schemes. In the zero-shot scheme, classification performance is moderate, reflecting the model's limitations in understanding the nuances of the marine tourism domain without the support of contextual examples. In contrast, the application of some-shot learning shows significant improvements in precision, recall, and F1 scores across all sentiment classes.

These findings confirm that the inclusion of a small number of labeled examples can help the model distinguish boundaries between sentiment classes more consistently, especially in multifaceted and ambiguous data sets.

The XAI-based analysis strengthens the quantitative results by demonstrating that the resulting sentiment prediction model is supported by semantically relevant keyword contributions. The alignment between keyword contributions, sentiment polarity, and review context indicates that the model does not operate as a black box, but rather generates traceable and operational decisions. This provides significant added value for marine tourism analytics applications, where analysis results are often used as a basis for policy formulation or service improvements.

Furthermore, cross-platform analysis shows that the proposed approach is able to adapt to the different characteristics of reviews on Google Maps, TripAdvisor, Instagram, and YouTube. Despite differences in language style, text length, and intensity of sentiment expression, the model still produces consistent sentiment patterns. These findings demonstrate that LLMs have good generalization capabilities to handle the

heterogeneity of user-generated content in the marine tourism context.

IV. CONCLUSION

This study proposes a marine tourism sentiment analysis framework based on a Large Language Model (LLM) with an Aspect-Based Sentiment Analysis (ABSA) approach integrated with Explainable Artificial Intelligence (XAI). The results show that the proposed approach is able to identify key aspects in marine tourism reviews and effectively classify user sentiments on multi-aspect and informal user-generated content data. Quantitatively, the evaluation model shows a significant performance difference between the zero-shot and some-shot learning schemes. In the zero-shot scheme, the model achieved an accuracy of 0.33 with a macro-average F1-score of 0.33, reflecting the model's limitations in understanding the context of the marine tourism domain without the support of labeled examples. In contrast, the application of some-shot learning substantially improved the classification performance with an accuracy of 0.87 and a macro-averaged F1-score of 0.87, indicating that the inclusion of a small number of contextual examples in the prompt significantly improves the accuracy and consistency of sentiment predictions. Furthermore, the XAI integration results show that the predicted sentiment generated by the model is supported by the contribution of semantically relevant keywords and phrase evidence. The congruence between predicted sentiment, extracted aspects, and linguistic evidence enhances the model's transparency and interpretability, preventing the analysis from being a black box. These findings confirm that the proposed framework not only performs well but is also reliable as an analytical tool to support decision-making in marine tourism destination management.

As a further development direction, this research could address increasing the scale and diversity of manual annotation data to strengthen quantitative evaluation and generalize the results. Future research could also explore more adaptive prompt engineering strategies, including varying the number and structure of examples in some-shot schemes, to optimize model performance on infrequent aspects.

ACKNOWLEDGMENT

The authors sincerely acknowledge the Doctoral Scholarship Program for Indonesian Lecturers, Center for Higher Education Funding and Assessment, and Ministry of Higher Education, Science and Technology of the Republic of Indonesia, for its continuous financial support and academic facilitation that significantly contributed to the accomplishment of this study.

REFERENCES

- [1] M. Carvache-Franco, A.G. Viquez-Paniagua, W. Carvache-Franco, A. Pérez-Orozco, and O. Carvache-Franco, "Segmentation by Motivation in Sustainable Coastal and Marine Destinations: A Study in Jacó, Costa Rica," *Sustain.*, vol. 14, no. 14, pp. 1–14, 2022, doi: 10.3390/su14148830.
- [2] KM Aboalghanam, SF AlFraihat, and S. Tarabieh, "The Impact of User-Generated Content on Tourist Visit Intention: The Mediating Role of Destination Image," *Adm. Sci.*, vol. 15, no. 4, 2025, doi: 10.3390/admsci15040117.
- [3] S. Sujatmiko, DP Ar, A. Hamdat, and KN Salam, "User-Generated Content (UGC) and Its Impact on Tourism Marketing: A Systematic

- Literature Review,” *Golden Ratio Mapp. Idea Lit. Format*, vol. 5, no. 2, pp. 97–105, 2025, doi: 10.52970/grmilf.v5i2.1491.
- [4] EC Ramos and CMQ Ramos, “User-Generated Content and Its Impact on Tourism Product Purchase Intention: A Comparative Analysis of Millennials and Centennials on TikTok,” *Futur. Internet*, vol. 17, no. 3, pp. 1–21, 2025, doi: 10.3390/fi17030105.
- [5] HK Tunjungsari, V. Willyanto, WC Hoo, and CW Wolor, “How User-Generated Content Influences Tourist Loyalty Behavior at Cultural Heritage Sites in Indonesia,” *Int. J. Sustain. Dev. Plan.*, vol. 20, no. 3, pp. 1243–1252, 2025, doi: 10.18280/ijstdp.200329.
- [6] NMS Iswari and N. Afriliana, “Improving Aspect-Based Sentiment Analysis in Tourism Reviews Through Hybrid Data Augmentation,” *J. Appl. Data Sci.*, vol. 6, no. 3, pp. 2192–2206, 2025, doi: 10.47738/jads.v6i3.842.
- [7] HA Laudza, J. Tegar, A. Putra, MD Attsaqif, and IK Setiawan, “Sentiment Analysis and Topic Modeling of Tourist Reviews on Lawang Sewu Cultural Heritage Destination,” *Int. J. Adv. Sci. Res. Eng.*, vol. 11, no. 9, pp. 38–48, 2025, doi: 10.31695/IJASRE.2025.9.3.
- [8] DC Rahmadani, S. Khomsah, and MY Fathoni, “Analysis of Tourist Emotions Using the Lexicon Text Analysis Method,” *J. Tech. Inform. and Kak. Inf.*, vol. 10, no. 1, pp. 13–27, 2024, doi: 10.28932/jutisi.v10i1.6690.
- [9] NW Utami, IN Purnama, and IMA Prayoga, “Bali Tourism Sentiment Analysis System Using Naive Bayes Algorithm and Web Framework,” *J. Info Sains*, vol. 14, no. 03, pp. 275–282, 2024, doi: 10.54209/infosains.v14i03.
- [10] NQ Fitriyah et al., “Sentiment Analysis of Reviews about Ijen Crater Using the Naive Bayes Classification Method,” in *Beyond Technology Summit on Informatics (BTS-I2C) 2024 International Conference*, 2024, pp. 509–514. doi: 10.1109/BTS-I2C63534.2024.10941799.
- [11] ESJ Atmadji, Y. Wabula, HT Karsanti, and K. Kristopher, “Use of Naive Bayes Classification in Sentiment Analysis in Indonesia's Super Priority Tourism Destinations Based on User Reviews,” *J. Sosioteknologi*, vol. 24, no. 2, pp. 226–239, 2025, doi: 10.5614/sostek.itbj.2025.24.2.7.
- [12] YA Singgalen, SY Wahyuningtyas, YE Widodo, MNA Dasra, and RW Setiawan, “Travel Vlog Review: Support Vector Machine Performance in Sentiment Classification,” *Ing. des Syst. d'Information*, vol. 30, no. 1, pp. 101–110, 2025, doi: 10.18280/isi.300109.
- [13] Helen and R. Kurniawan, “Tourism Destination Sentiment Analysis Using Support Vector Machine Algorithm on Twitter Posts,” in *International Conference on Information Management and Technology (ICIMTech) 2023*, 2023, pp. 316–321. doi: 10.1109/ICIMTech59029.2023.10277866.
- [14] S. Tabinda Kokab, S. Asghar, and S. Naz, “Transformer-based deep learning model for sentiment analysis of social media data,” *Array*, vol. 14, no. February, p. 100157, 2022, doi: 10.1016/j.array.2022.100157.
- [15] V. Balakrishnan, Z. Shi, C. L. Law, R. Lim, L. L. Teh, and Y. Fan, “Deep learning approaches in predicting product sentiment ratings: a comparative analysis,” *J. Supercomput.*, vol. 78, no. 5, pp. 7206–7226, 2022, doi: 10.1007/s11227-021-04169-6.
- [16] AS Talaat, “A sentiment analysis classification system using a hybrid BERT model,” *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00781-w.
- [17] ZM Zohreh Madhoushi, AR Hamdan, and S. Zainudin, “Aspect-Based Sentiment Analysis Methods in Recent Years,” *Asia-Pacific J. Inf. Technol. Multimed.*, vol. 08, no. 01, pp. 79–96, 2019, doi: 10.17576/apjitm-2019-0801-07.
- [18] R. Jayanto, R. Kusumaningrum, and A. Wibowo, “Aspect-based sentiment analysis for hotel reviews using an improved long short-term memory model,” *Int. J. Adv. Intell. Informatics*, vol. 8, no. 3, pp. 391–403, 2022, doi: 10.26555/ijain.v8i3.691.
- [19] W. Ahmad, HU Khan, FK Alarfaj, and M. Alreshoodi, “Aspect-Based Sentiment Analysis: A Comprehensive Review and Open Research Challenges,” *IEEE Access*, vol. 13, no. March, pp. 65138–65182, 2025, doi: 10.1109/ACCESS.2025.3555744.
- [20] DI Afidah, PD Anggraeni, M. Rizki, AB Setiawan, and SF Handayani, “Aspect-Based Sentiment Analysis for Review of Indonesian Tourism Attractiveness Using Bidirectional Long Short-Term Memory,” *JUITA J. Inform.*, vol. 11, no. 1, p. 27, 2023, doi: 10.30595/juita.v11i1.15341.
- [21] Y. Chai, H. Xie, and J. S. Qin, “Text Data Augmentation for Large Language Models: A Comprehensive Survey of Methods, Challenges, and Opportunities,” 2025, doi: 10.1007/s10462-025-11405-5.
- [22] Youngjin Chae and Thomas Davidson, “Large Language Models for Text Classification: From Zero-Shot Learning to Instruction Adaptation,” *Sociol. Methods Res.*, p. 00491241251325243, Apr. 2025, doi: 10.1177/00491241251325243.
- [23] J. Rui, Y. Xu, C. Cai, and X. Li, “Leveraging large language models for tourism research based on the 5D framework: Collaborative analysis of tourist sentiment and spatial features,” *Tour. Manag.*, vol. 108, no. December 2024, 2025, doi: 10.1016/j.tourman.2024.105115.
- [24] S. Ali et al., “Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence,” *Inf. Fusion*, vol. 99, no. April, p. 101805, 2023, doi: 10.1016/j.inffus.2023.101805.
- [25] Arunraju Chinnaraju, “Explainable Artificial Intelligence (XAI) for trustworthy and transparent decision making: A theoretical framework for AI interpretability,” *World J. Adv. Eng. Technol. Sci.*, vol. 14, no. 3, pp. 170–207, 2025, doi: 10.30574/wjaets.2025.14.3.0106.
- [26] DE Mathew, DU Ebem, AC Ikegwu, PE Ukeoma, and NF Dibiaezue, “Emerging Techniques in Explainable Artificial Intelligence to Improve Human Interpretation and Understanding of AI Models,” *Neural Process. Lett.*, vol. 57, no. 1, pp. 1–32, 2025, doi: 10.1007/s11063-025-11732-2.
- [27] H. Ruan, X. Li, X. Li, H. Jiang, and Y. Li, “Less Data Learning for BT Aspect-Based Sentiment Analysis - Advances in Natural Computing, Fuzzy Systems and Knowledge Discovery,” N. Xiong, M. Li, K. Li, Z. Xiao, L. Liao, and L. Wang, Eds., Cham: Springer International Publishing, 2023, pp. 1146–1157.
- [28] Z. Wang, Q. Xie, and R. Xia, “A Simple yet Effective Framework for Few-Shot Aspect-Based Sentiment Analysis,” *SIGIR 2023 - Proc. 46th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, no. January, pp. 1765–1770, 2023, doi: 10.1145/3539618.3591940.