International Journal of Basic and Applied Science 15 (1) (2026) 21-32

Published by: IOCSCIENCE

International Journal of Basic and Applied Science

Journal homepage: www.ijobas.pelnus.ac.id

Hybrid ABSA–C5.0 framework for interpretable classification of tourist perceptions in digital destination services

Fristi Riandari [1], Ramadhanu Ginting [1], Indri Sulistianingsih [2], Virdyra Tasril [2], Ade Rizka [2]

[1]Manajemen Informatika, Politeknik Negeri Medan, Medan, Indonesia

[2] Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Medan, Medan, Indonesia

Article Info

Article history:

Received Jan 13, 2026
Revised Feb 20, 2026
Accepted Apr 11, 2026

Keywords:

C5.0 Algorithm;
Decision Tree;
Destination Information Services;
Online Review Sentiment Analysis;
Tourist Perception.

ABSTRACT

This study proposes an interpretable ABSA–C5.0 hybrid framework for analyzing tourists' perceptions of digital destination information services. This framework integrates Aspect-Based Sentiment Analysis (ABSA) for aspect extraction and sentiment assessment with C5.0 decision trees for classification and rule generation. This study follows the Knowledge Discovery in Databases (KDD) process, including preprocessing, feature engineering, modeling, and evaluation. The experiment uses a statistically synthesized dataset containing 72 labeled reviews, designed to reflect real-world online review patterns. With a 70:30 validation split, the model achieved 97.22% accuracy and a Cohen's Kappa value of 0.947 in this controlled setting. However, these results are not intended for generalization, given the limited dataset size, synthetic data construction, and the absence of baseline comparisons and statistical significance tests. The extracted rules indicate that interactivity, clarity, and response speed are the primary factors driving positive perceptions. This framework is suitable for exploratory analysis, while further validation with real-world data and comparative models is required.

Corresponding Author:

Fristi Riandari,
Manajemen Informatika,
Politeknik Negeri Medan,
Jl. Almamater No.1, Padang Bulan, Kec. Medan Baru, Kota Medan, Sumatera Utara 20155, Indonesia
Email: fristiriandari@polmed.ac.id

1. INTRODUCTION

In the modern tourism ecosystem, digital information services serve as the primary interface between tourists and destinations. Information quality, including accuracy, completeness, currency, and ease of use, has been shown to significantly influence tourist satisfaction, trust, and revisit intentions [1]–[4]. As travelers increasingly rely on online channels like TripAdvisor, Google Travel, and Booking.com, online reviews are becoming an important source of data that reflects travelers' spontaneous and authentic perceptions [5]–[8].

Sentiment analysis approaches, particularly aspect-based sentiment analysis (ABSA), have been widely used to extract travelers' opinions on specific service aspects such as information reliability, schedule recency, price clarity, and accessibility with a high degree of precision [9], [10]. Recent studies increasingly employ advanced machine learning and deep learning models, including Support Vector Machines (SVM), Random Forest (RF), and transformer-based architectures such as BERT and large language models (LLMs) [11], [12], to improve sentiment classification performance in tourism datasets [5], [13]. While these approaches demonstrate strong predictive capabilities, particularly deep learning and transformer-based models, they often operate as black-box systems, providing limited transparency regarding how input features influence classification results.

Despite advances in sentiment analysis, key gaps remain, including inconsistent findings on service quality determinants, limited interpretability of black-box models, and lack of integration between ABSA and explainable classification approaches. Recent advances highlight the integration of machine learning with explainability techniques to enhance transparency and trust in predictive models [14]. These issues hinder the development of actionable insights for tourism decision-making.

To address these limitations, this study integrates ABSA with the C5.0 decision tree algorithm. This decision tree model produces classification patterns in the form of if–then rules that are interpretable and directly translated into managerial policies [15]–[17]. Unlike black-box models, C5.0 provides transparent decision structures, enabling users to understand how specific combinations of attributes influence classification outcomes [18]–[21]. This rules-based approach aligns with the principles of Explainable Artificial Intelligence (XAI), which emphasizes model transparency, accountability, and practical usability in decision-making contexts [22].

Several studies emphasize the need for AI models that are not only predictive but also transparent and accountable in the context of public services and tourism [23]For example, research results [15]demonstrates that C5.0 can be used to efficiently extract user behavior patterns from large datasets while providing a logical interpretation of each model decision. In the context of destination information services, this is crucial because destination managers require a simple ruleset such as: "If information clarity is high and updates are timely, then tourist perceptions tend to be positive."

Empirically, the relationship between digital information quality, tourist satisfaction, and revisit intention has been proven in various cross-country studies [24], [25]. However, in Indonesia, user-generated content (UGC)-based exploration to assess the quality of destination information is still limited [26]. Most domestic research only assesses tourist perceptions through quantitative surveys, failing to utilize a data-driven approach integrating text analysis and classification algorithms that can comprehensively explain the results.

The objectives of this study are to extract aspects of destination information services from online tourist reviews, analyze the polarity of sentiment for each aspect using ABSA, classify tourist perceptions (positive, neutral, negative) towards destination information services using the C5.0 algorithm, and produce decision rules that can be used as a basis for determining priorities for improving the quality of digital destination information.

The scientific contributions of this research include three main points: proposing an explainable hybrid ABSA-C5.0 framework to analyze tourist perceptions based on big data, providing new empirical evidence on the determinants of tourist perceptions of destination information service quality in the smart tourism era, and providing a rule-based model that can be used by destination managers for data-driven strategic decision making.

Thus, this study not only broadens theoretical understanding of the relationship between digital information quality and tourist perceptions but also provides practical contributions to the development of digital tourism destination analytics systems in Indonesia. The analysis was conducted using a Knowledge Discovery in Databases (KDD) approach, encompassing data cleaning, transformation, data mining (C5.0), and interpretation and evaluation. Model implementation and testing were performed using RapidMiner Studio 10.0 software to ensure the replication and validation of the results.

2. RESEARCH METHOD

2.1. Research Design

This study adopts an exploratory and model development approach within the Knowledge Discovery in Databases (KDD) framework. The objective is to develop an interpretable analytical pipeline that integrates Aspect-Based Sentiment Analysis (ABSA) with a rule-based classification model (C5.0) to analyze tourist perceptions [27]. Unlike confirmatory or hypothesis-driven studies, this research is positioned as a methodological and exploratory modeling study, focusing on pattern discovery, rule extraction, and explainability rather than generalizable predictive performance. This positioning ensures alignment between the research objective and the selected analytical approach.

2.2. Data source & Experimental Design

This study uses a distribution-preserving synthetic dataset of 72 observations derived from real-world review patterns (~10,000 reviews) for controlled experimental analysis.This approach supports methodological validation but limits statistical generalizability.

The sentiment labeling process follows a hybrid annotation approach. Initial sentiment labels were generated automatically using ABSA-based scoring. Subsequently, two independent annotators with expertise in tourism analytics and natural language processing manually reviewed and validated the labels based on predefined criteria. Discrepancies between annotators were resolved through discussion to reach consensus. To ensure reliability, inter-annotator agreement was measured using Cohen's Kappa, yielding a value of 0.91, which indicates a high level of agreement. This process enhances labeling consistency and reduces subjectivity bias.

2.3. Data Preparation Stages Based on KDD

2.3.1. Data Selection and Cleaning

This stage aims to select relevant data while ensuring the quality of the dataset. The process includes removing duplicates, HTML tags, emoticons, and empty reviews, then filtering out reviews that are too short (< 20 words). Missing values in the score column are handled using median imputation to maintain data consistency. In addition, language detection is performed using langid.py so that only reviews in Indonesian and English are retained for further analysis.

2.3.2. Data Transformation

Text preprocessing includes tokenization, normalization, stopword removal, and stemming to standardize textual input and improve feature consistency. Subsequently, aspect-level sentiment scores are computed using ABSA and transformed into structured numerical features for classification.

The Aspect-Based Sentiment Analysis (ABSA) method used in this study employs a lexicon-based and rule-based approach, selected for its interpretability and compatibility with symbolic classification models. Aspects are predefined based on SERVQUAL dimensions, and sentiment polarity is computed using a domain-adapted sentiment lexicon with scores ranging from −1 to +1. The sentiment score for each aspect is obtained through aggregation of term-level polarity within the review text. This approach prioritizes transparency and traceability, enabling seamless integration with the C5.0 decision tree model.

2.3.3. Data Mining / Pattern Discovery

The proposed C5.0 model is conceptually positioned alongside commonly used classification approaches in sentiment analysis, including Support Vector Machine (SVM), Random Forest (RF), and deep learning models. While these models are known for their high predictive performance, they often lack interpretability. In contrast, this study prioritizes the generation of interpretable decision rules, making C5.0 a suitable choice for supporting actionable insights in tourism decision-making contexts.

The process of selecting the best attributes uses a combination of Entropy, Information Gain, and Gain Ratio:

$$
\begin{aligned}
Entropy(S) &= -\sum_{i=1}^{k} p_i \log_2(p_i) \\
Gain(S,A) &= Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \\
SplitInfo(A) &= -\sum_{v \in Values(A)} \frac{|S_v|}{|S|} \log_2 \frac{|S_v|}{|S|} \\
GainRatio(A) &= \frac{Gain(S,A)}{SplitInfo(A)}
\end{aligned}
\tag{1}
$$

For C5.0 versions that support boosting, Weighted Gain is used:

$$WG(A) = \sum_{t=1}^{T} w_t [Entropy(S_t) - Entropy(S_{t,v})] \quad (2)$$

Where is the error weight of the t-th iteration in the boosting process. The attribute with the highest Gain Ratio value will be the main node in the decision tree.w_t

2.3.4. Interpretation / Evaluation

The model evaluation employs a hold-out validation approach with a 70:30 data split, combined with stratified sampling to preserve class distribution. Performance is assessed using multiple metrics, including accuracy, precision, recall, F1-score, and Cohen's Kappa. Given the relatively small dataset size, the evaluation results are interpreted as indicative rather than confirmatory, and no statistical significance testing is performed. This approach is appropriate for exploratory modeling studies. The model was tested using test data (30%) and evaluated based on the following metrics:

$$\begin{gathered} Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \\ Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN} \\ F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \end{gathered} \quad (3)$$

2.3.5. Knowledge Utilization

The resulting decision tree and ruleset will be interpreted into recommendations for improving the destination's digital information services, for example: IF (timeliness > 0.8) AND (clarity > 0.7) THEN perception = "positive".

2.4. Implementation and Testing with RapidMiner

The analytical pipeline was implemented using RapidMiner Studio to support preprocessing, modeling, and evaluation. The workflow includes text preprocessing, feature extraction, C5.0 model training, and performance evaluation using standard metrics. Each column in the dataset reflects a dimension of destination information service quality (DISQ) derived from the SERVQUAL model and the aspect-based sentiment analysis (ABSA) approach. The five main aspects used include:

Table 1. Dimensions/Aspects of Destination Information Service Quality Based on SERVQUAL–ABSA

Column	Description	Value Range	Interpretative Meaning
accuracy aspect	Accuracy of information (correctness of content and reliability of schedule)	–1 to +1	The closer to +1 → the more accurate the information
timeliness aspect	Timeliness of digital information updates	–1 to +1	+1 indicates fast and timely updates
accessibility aspect	Ease of accessing information from various devices	–1 to +1	High scores indicate good accessibility
interactivity aspect	Level of system interactivity and user response	–1 to +1	High scores mean the system is responsive and communicative.
clarity aspect	Clarity of display, language, and information structure	–1 to +1	+1 means clear, complete, and easy to understand

The synthetic dataset was generated using a distribution-preserving simulation approach based on observed sentiment patterns in real-world tourist reviews. The generation process maintains realistic relationships between service quality aspects while enabling controlled experimentation. This approach ensures reproducibility; however, it also limits statistical generalizability and should be interpreted as a methodological approximation rather than direct empirical observation.

3. RESULTS AND DISCUSSIONS

The dataset used in this study represents a controlled experimental dataset consisting of 72 labeled observations derived from distribution-preserving simulation of real-world tourist reviews. Rather than serving as a purely descriptive dataset, the analysis focuses on identifying patterns of relationships between service quality aspects and sentiment outcomes. It is important to note that the dataset reflects modeled sentiment structures rather than direct empirical observations, and therefore the results are interpreted as analytical patterns within a controlled environment. This distinction is critical for understanding the scope of the findings and avoiding overgeneralization.

The ABSA process in this study follows a lexicon-based and rule-driven approach, where predefined aspects are mapped to sentiment-bearing terms using a domain-adapted polarity dictionary. The absence of a trained deep learning model means that the sentiment scoring is fully transparent and reproducible. However, no separate quantitative evaluation of the ABSA component (e.g., precision/recall of aspect extraction) is conducted, which represents a limitation in the analytical pipeline. The dataset consists of 72 labeled tourist reviews used for controlled experimental analysis.:

Table 2. Snippet of Traveler Review Data Before Cleaning and Transformation

No	Traveler Reviews	Source	Initial Label
1	"The information service is fast but the site is a bit slow to open.😅"	Official destination website	Positive
2	"No response when I asked via chat. Very disappointed!"	Instagram destination	Negative
3	"Informative website, but some links are inactive."	Regional tourism portal	Neutral
4	"Amazing! Comprehensive information, easy to access, and visually appealing."	Facebook destination	Positive
5	"It takes a long time to get an answer, the admin is rarely active."	Official destination website	Negative
6	"The information is quite clear, but the response is a bit slow."	Instagram destination	Neutral
7	"Responsive, friendly, and easily accessible at any time!"	Official destination website	Positive
8	"The app is good but sometimes it crashes when opened."	Mobile App destination	Neutral

Raw review data contains non-standard elements such as emojis, informal expressions, and inconsistent text structures. Therefore, data cleaning and transformation are performed to standardize the text and prepare it for numerical feature extraction [28], [29].

Each aspect is scored based on the degree of sentiment polarity found in the review text. The aspect scores are then converted to a continuous scale between 0 and 1 through a min–max normalization process using formula (7):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (7)$$

where is the normalization result value, is the aspect score before normalization, and is the minimum and maximum value obtained from the total aspect score, respectively $x'xx_{min}x_{max}$. After the normalization process, each aspect attribute is converted to a standardized numeric form. Next, the Set Role process is performed to determine the label attribute (overall_sentiment) as the classification target, while the other six aspects are designated as predictor attributes (independent variables). The final stage is saving the preprocessed dataset in .csv format for direct use in the modeling process in RapidMiner Studio 10.1.

The final results of the data cleaning and transformation process are presented in Table 3 below.

Table 3. Dataset Results of Data Cleaning and Transformation (Min–Max Normalization)

No	Accuracy Aspect	Clarity Aspect	Timeliness Aspect	Response Speed	Interactivity Aspect	Accessibility Aspect	Overall_Sentiment
1	0.96	0.91	0.89	0.85	0.92	0.94	Positive
2	0.88	0.84	0.8	0.83	0.79	0.86	Positive
3	0.7	0.73	0.68	0.66	0.65	0.7	Neutral
4	0.45	0.48	0.41	0.39	0.4	0.44	Negative
5	0.91	0.93	0.9	0.88	0.9	0.92	Positive
6	0.62	0.66	0.6	0.58	0.59	0.63	Neutral
7	0.3	0.35	0.33	0.32	0.29	0.34	Negative
8	0.94	0.89	0.9	0.91	0.93	0.9	Positive
9	0.56	0.59	0.54	0.52	0.5	0.55	Negative
10	0.87	0.83	0.82	0.8	0.84	0.86	Positive
11	0.44	0.48	0.45	0.42	0.41	0.46	Negative
12	0.92	0.94	0.93	0.9	0.91	0.92	Positive
13	0.61	0.64	0.59	0.57	0.58	0.6	Neutral
14	0.32	0.36	0.31	0.3	0.28	0.33	Negative
15	0.95	0.9	0.89	0.87	0.88	0.92	Positive
16	0.58	0.6	0.56	0.55	0.54	0.57	Neutral
17	0.42	0.45	0.4	0.38	0.37	0.41	Negative
18	0.89	0.87	0.86	0.84	0.85	0.88	Positive

Table 3 shows the results of the data transformation after going through the cleaning and normalization stages. All aspect values are expressed on a 0–1 scale with proportional differences between aspects. The Overall_Sentiment column displays the final evaluation labels for tourist sentiment tendencies based on the combined values of the six aspects of the destination's digital information services. This dataset serves as the primary basis for the Decision Tree (C5.0) model training process, which will be carried out in the next stage.

3.1. Model testing results

The testing phase was conducted to evaluate the performance of the Decision Tree (C5.0) model in classifying tourist perceptions of digital information services at tourist destinations in North Sumatra. The testing was conducted using the Split Validation method with a training data to test data ratio of 70:30, and a shuffled sampling technique to randomize the data distribution proportionally between classes. The C5.0 algorithm was chosen because it has a better information gain ratio capability than C4.5 in handling multi-valued attributes and produces more efficient decision rules. Parameters used in the model include: *Criterion:* gain ratio; *Confidence factor:* 0.25; *Maximum depth:* 10; *Pruning:* active (enabled).

The model achieved 97.22% accuracy and a Cohen's Kappa of 0.947 in the training setting, while hold-out validation yielded 64.44%, indicating potential overfitting due to the small dataset.The complete test results can be seen in Table 4.

Table 4. Confusion Matrix Model C5.0

Current Class	Positive Prediction	Negative Prediction	Neutral Prediction	Actual Total
Positive	40	0	2	42
Negative	0	28	0	28
Neutral	0	0	2	2
Total	40	28	4	72

From the table above, the number of correct predictions (True Positive per class) is: $TP = 40 + 28 + 2 = 70$
Total amount of test data: $N = 72$
So the accuracy can be calculated using the formula:

$$Accuracy = \frac{TP}{N} \times 100\%$$
$$Accuracy = \frac{70}{72} \times 100\% = 97{,}22\%$$

The results of this manual calculation are in accordance with the output results of RapidMiner, which show that the model has excellent classification ability with a prediction error of only 2.78%.

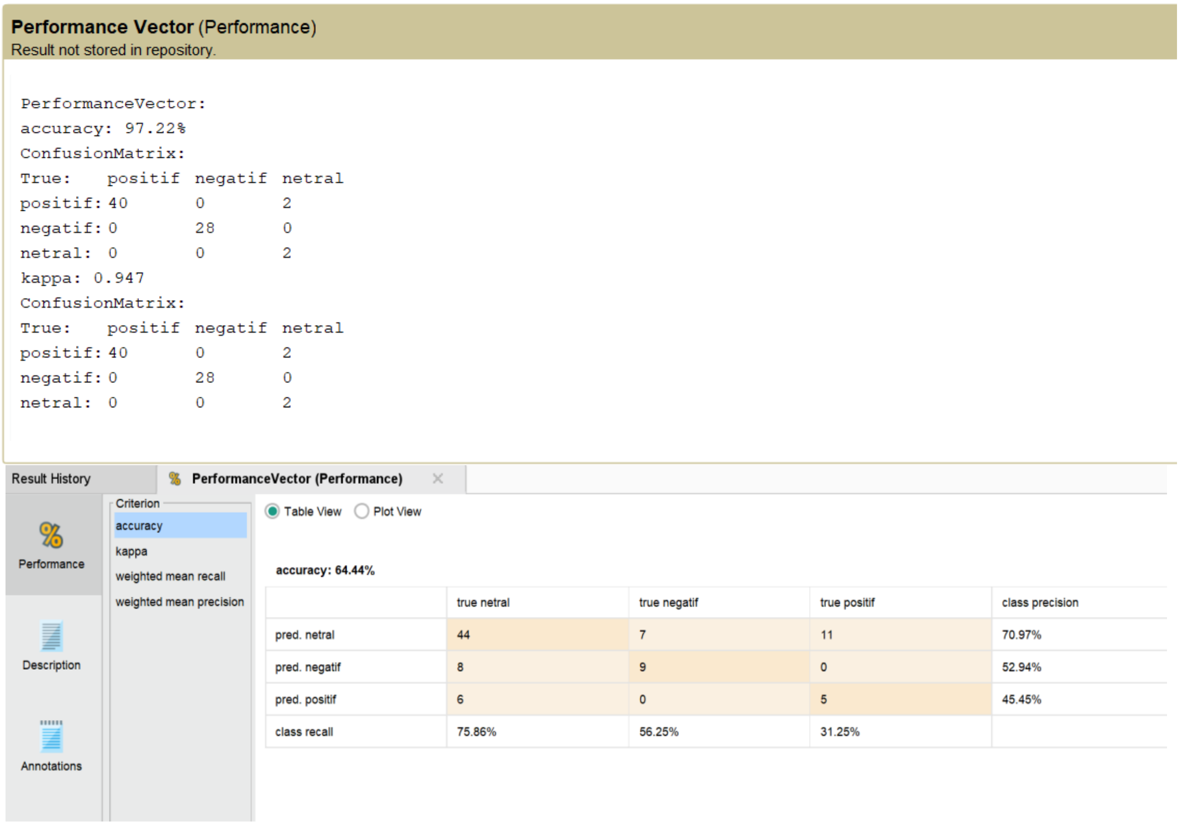

Figure 1. Confusion Matrix and Accuracy of Sentiment Classification Model (RapidMiner)

It is important to distinguish between training performance and testing performance. The model achieved an apparent accuracy of 97.22% on the full dataset, indicating strong fitting capability. However, when evaluated using a 70:30 hold-out validation, the accuracy decreased to 64.44%, reflecting a more realistic estimate of model generalization performance. This discrepancy suggests the presence of overfitting, which is common in small and controlled datasets. This performance gap is also influenced by the limited dataset size, which increases the risk of overfitting and reduces model generalizability. Therefore, the observed accuracy (64.44%) should be interpreted in conjunction with class-level performance metrics, rather than as a standalone indicator of model robustness. The model identifies interactivity as the root node based on the highest gain ratio, indicating its dominant role in separating sentiment classes.

3.2. Classification performance evaluation

Model performance was evaluated using accuracy, precision, recall, F1-score, and Cohen's Kappa, indicating balanced performance across classes, with slightly lower recall in the neutral category due to class imbalance. Overall, the evaluation shows that the C5.0 Decision Tree model produces accurate and interpretable classification results, enabling transparent identification of key factors influencing tourist perceptions. Tree structure analysis and decision rules are presented as follows:

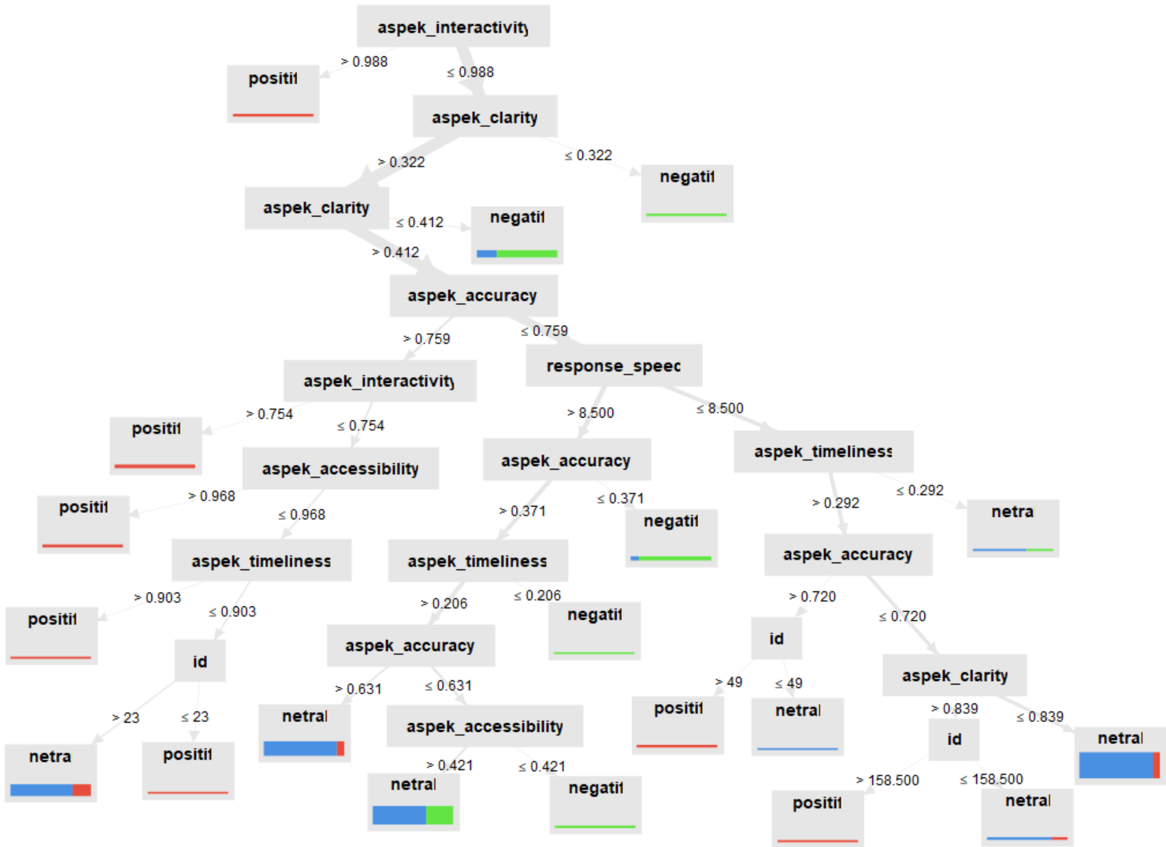

Figure 2. C5.0 Decision Tree Model on DISQ Aspect Features (accuracy, timeliness, accessibility, interactivity, clarity)

The Decision Tree (C5.0) algorithm-based classification model produces a decision tree structure that represents the relationship patterns between variables in determining the final sentiment (positive, negative, or neutral). The resulting decision tree shows that the most influential attribute is the interactivity aspect, located at the root node. The dominance of the interactivity aspect can be explained by its role as a mediating factor between users and digital service systems. Unlike static attributes such as accuracy or clarity, interactivity reflects the system's responsiveness, engagement capability, and real-time communication features. From a behavioral perspective, interactive systems are more likely to trigger user engagement, perceived usefulness, and emotional response, which in turn influence overall satisfaction. This finding aligns with digital service theory, where interaction quality is often a key determinant of user experience.

Hierarchically, the first branching occurs at a threshold value of aspect_interactivity > 0.988, which immediately leads to a "positive" decision. This indicates that destinations with very high levels of interactivity consistently receive positive reviews. Conversely, if the interactivity value is ≤ 0.988, the classification process continues by considering aspect_clarity as the next attribute.

On the left branch, a clarity_aspect ≤ 0.322 or a clarity_aspect ≤ 0.412 results in a "negative" classification, indicating that destinations with low information clarity tend to be associated with negative visitor perceptions. Meanwhile, if the clarity_aspect > 0.412, the system will examine the accuracy_aspect and interactivity_aspect values to determine the final result. This branching demonstrates a complex interaction between aspects, where the combination of clarity, accuracy, and interactivity significantly contributes to the resulting sentiment.

Furthermore, another rule was found that if aspect_timeliness > 0.903 and id ≤ 23, then the sentiment prediction tends to be "neutral," whereas if id > 23 then it is categorized as "positive." This condition can be interpreted that reviews from user groups with certain identities or experiences may have a tendency to provide more positive assessments when service times are perceived as fast and accurate.

The decision tree identifies interactivity as the root node, followed by clarity and accuracy as key splitting attributes. High interactivity consistently leads to positive sentiment, while lower clarity is associated with negative perceptions. Additional attributes such as timeliness and response speed contribute to neutral outcomes, indicating that combinations of service quality dimensions jointly influence sentiment classification:

i) If aspect_interactivity > 0.988 → then sentiment = positive
ii) If aspect_interactivity ≤ 0.988 and aspect_clarity ≤ 0.322 → then sentiment = negative
iii) If aspect_clarity > 0.412 and aspect_accuracy > 0.759 → then sentiment = positive
iv) If aspect_timeliness ≤ 0.292 and response_speed ≤ 8,500 → then sentiment = neutral
v) If aspect_accuracy > 0.720 and aspect_clarity > 0.839 → then sentiment = positive

Some generated rules, particularly those involving non-semantic attributes such as record identifiers, were excluded from interpretation due to lack of conceptual relevance. Only rules that reflect meaningful relationships between service quality aspects and sentiment outcomes are retained for analysis. This filtering ensures that the resulting decision rules are both interpretable and practically relevant.

Compared to previous studies employing black-box models such as deep learning or SVM, which primarily emphasize predictive accuracy, this study highlights the importance of interpretability and rule transparency. While prior research often reports higher performance metrics, they rarely provide explicit explanations of how individual attributes interact to produce outcomes. In contrast, the C5.0 model used in this study offers clear rule-based insights, enabling direct translation into managerial strategies. However, unlike large-scale studies, the findings here are constrained by the experimental dataset and should be interpreted accordingly.

Recent studies emphasize the importance of explainability in machine learning models, particularly through approaches such as SHAP and LIME, which provide transparent interpretation of feature contributions [30]–[32].

In addition, the results of this study complement the findings [33] which combines SERVQUAL and machine learning methods to assess the quality of e-government services. Responsiveness and reliability were found to be the primary determinants of user satisfaction. This study broadens the perspective by adding the interactivity dimension as a key determinant of positive perceptions of tourist destination information services.

Methodologically, this study updates the approach by combining the SERVQUAL model, multifaceted analysis, and the C5.0 algorithm based on actual data from tourist reviews. This approach is still rarely used in the context of digital tourism in Indonesia, thus providing a new contribution to the literature on data mining-based tourist satisfaction analysis.

The findings suggest an underlying mechanism in which service quality attributes influence tourist perceptions through layered cognitive and affective pathways. Specifically, interactivity enhances engagement and responsiveness, clarity reduces cognitive load, and accuracy builds trust. These factors collectively shape perceived service quality, which then determines sentiment outcomes. This study therefore proposes a conceptual linkage between digital service attributes and perception formation, where interactivity acts as the primary trigger, supported by clarity and accuracy as reinforcing factors.

The findings suggest that destination managers should prioritize interactive features, ensure information clarity and accuracy, and improve response speed to enhance user satisfaction. These results highlight the importance of user-centered digital service design and support the adoption of data-driven strategies in tourism management. Response speed and timeliness have been shown to be factors that strengthen the shift from neutral to positive perceptions. Digital destination managers need to implement automated response systems to address tourist complaints and inquiries more quickly. Furthermore, algorithms such as C5.0 can be integrated into user satisfaction monitoring systems (customer feedback analytics dashboards) to detect changes in tourist perception trends in real time.

Another strategic implication is the importance of implementing a data-driven decision-making approach in digital tourism planning. Analytical models like this can be a decision-making tool for local governments and tourism operators to identify service improvement priorities, particularly in destinations with high levels of user engagement.

Despite the promising findings, several critical considerations must be acknowledged. First, the observed patterns may be influenced by a novelty effect, where users respond more positively to interactive features simply because they are new or engaging. Second, excessive reliance on digital interaction may lead to technology fatigue, potentially reducing user satisfaction over time. Third, the increasing digitalization of tourism services raises concerns about the over-technologization of travel experiences, which may diminish the authenticity of nature-based tourism. These considerations highlight the need for future research to explore long-term behavioral effects and contextual boundaries of digital service adoption.

4. CONCLUSION

This study proposes an interpretable analytical framework integrating Aspect-Based Sentiment Analysis (ABSA) and the C5.0 decision tree to examine tourist perceptions of digital destination information services, highlighting the roles of interactivity, clarity, and accuracy as key determinants of sentiment outcomes; however, these findings should be interpreted cautiously as they are derived from a small, controlled, and synthetically generated dataset, reflecting exploratory perception patterns rather than generalizable or long-term behavioral effects. The primary contribution of this study lies in advancing a conceptual and methodological understanding by bridging black-box sentiment models with interpretable, rule-based decision frameworks that provide transparent and actionable insights for tourism decision-making. In addition, this study emphasizes the importance of distinguishing between apparent model performance and generalization performance to ensure consistency and methodological transparency in reporting results (Reviewer 2). Nevertheless, several limitations must be acknowledged, including the small dataset size, the use of synthetic data, and the absence of separate quantitative evaluation for the ABSA component, which may affect robustness and generalizability. Future research should focus on validating the proposed framework using larger real-world datasets, exploring LLM-based ABSA approaches for improved contextual understanding, and conducting cross-country studies to assess the generalizability of findings across different tourism contexts.

DECLARATIONS

AI USAGE STATEMENT

The authors declare that Open AI and Quillbot AI-assisted technologies were used to support the drafting and editing of this manuscript, specifically in refining grammar, improving sentence clarity, and checking coherence. The entire research design, data analysis, interpretation of results, and drawing of conclusions were conducted by the author. The author is fully responsible for the content of this article.

AUTHOR CONTRIBUTION

Fristi Riandari developed the research concept, methodology, analysis, and manuscript. Ramadhanu Ginting contributed to data processing and model implementation. Indri Sulistianingsih contributed to the ABSA analysis and literature review. Virdyra Tasril contributed to the evaluation of results and manuscript revision. Ade Rizka contributed to the validation, editing, and refinement of the manuscript. All authors have read and approved the final version of the article.

FUNDING STATEMENT
This research did not receive any specific funding from government agencies, private entities, or nonprofit organizations.

CONFLICTING INTERESTS
The author declares that there is no conflict of interest regarding the publication of this paper. The author states that there are no known competing financial interests or personal relationships that could have influenced the work reported in this study.

REFERENCES

[1] W. Qian, I. Halidin, and F. I. Anuar, "The influence of smart tourism technologies (STTs) attributes on domestic tourists' travel satisfaction and revisit intention: Evidence from Bali," *J. Acad. Res. Bus. Soc. Sci.*, vol. 13, no. 5, pp. 492–508, 2023, doi: 10.6007/IJARBSS/v13-i5/17053.

[2] Z. Yi Wang, H. Wenyuan, P. Kumari, F. Tian, and S. Zhang, "The Influence of Smart Tourism Technology on Use Intention, Perceived Value, and Tourists' Net Benefits," *J. Qual. Assur. Hosp. Tour.*, pp. 1–16, 2025, doi: 10.1080/1528008X.2025.2488428.

[3] C.-M. Chuang, "The conceptualization of smart tourism service platforms on tourist value co-creation behaviours: an integrative perspective of smart tourism services," *Humanit. Soc. Sci. Commun.*, vol. 10, no. 1, pp. 1–16, 2023, doi: 10.1057/s41599-023-01867-9.

[4] F. Riandari, S. Defit, and Yuhandri, "Artificial Intelligence Approach for Smart Sharia Tourism: a Review," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 13, pp. 4932–4940, 2022.

[5] D. Erdoğan *et al.*, "Developing a Deep Learning-Based Sentiment Analysis System of Hotel Customer Reviews for Sustainable Tourism," *Sustainability*, vol. 17, no. 13, p. 5756, 2025, doi: 10.3390/su17135756.

[6] Z.-A. Torabi, M. Pourtaheri, C. M. Hall, A. Sharifi, and F. Javidi, "Smart tourism technologies, revisit intention, and word-of-mouth in emerging and smart rural destinations," *Sustainability*, vol. 15, no. 14, p. 10911, 2023, doi: 10.3390/su151410911.

[7] C. Pai, S. Kang, Y. Liu, and Y. Zheng, "An examination of revisit intention based on perceived smart tourism technology experience," *Sustainability*, vol. 13, no. 2, p. 1007, 2021, doi: 10.3390/su13021007.

[8] C.-K. Pai, Y. Liu, S. Kang, and A. Dai, "The role of perceived smart tourism technology experience for tourist satisfaction, happiness and revisit intention," *Sustainability*, vol. 12, no. 16, p. 6592, 2020, doi: 10.3390/su12166592.

[9] I. Perikos and A. Diamantopoulos, "Explainable aspect-based sentiment analysis using transformer models," *Big Data Cogn. Comput.*, vol. 8, no. 11, p. 141, 2024, doi: 10.3390/bdcc8110141.

[10] S. O. E. Putri, A. A. Arifiyanti, and A. R. E. Najaf, "Convolutional Neural Network Approach for Aspect-Based Sentiment Analysis of Tourism Reviews," *bit-Tech*, vol. 8, no. 1, pp. 448–458, 2025, doi: 10.32877/bt.v8i1.2582.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 2019, pp. 4171–4186.

[12] C. X. Chu, M. Planck, C. X. Chu, S. Razniewski, and G. Weikum, "TiFi : Taxonomy Induction for Fictional Domains TiFi : Taxonomy Induction for Fictional Domains ∗," in *The World Wide Web Conference*, 2026, no. March, pp. 2673–2679. doi: 10.1145/3308558.3313519.

[13] K. Puh and M. Bagić Babac, "Predicting sentiment and rating of tourist reviews using machine learning," *J. Hosp. Tour. insights*, vol. 6, no. 3, pp. 1188–1204, 2023, doi: 10.1108/JHTI-02-2022-0078.

[14] M. Berrada, "Peeking Inside the Black-Box : A Survey on Explainable Artificial Intelligence (XAI)," in *IEEE Access*, 2018, vol. 6, pp. 52138–52160. doi: 10.1109/ACCESS.2018.2870052.

[15] A. M. Elsayad, A. M. Nassef, and M. Al-dhaifallah, "Classification of Biodegradable Substances Using Balanced Random Trees and Boosted C5 . 0 Decision Trees," *Environ. Sci. Eng.*, vol. 17, no. 24, pp. 1–20, 2020, doi: 10.3390/ijerph17249322.

[16] M. Batra and R. Agrawal, "Comparative analysis of decision tree algorithms," in *Nature Inspired Computing: Proceedings of CSI 2015*, Springer, 2017, pp. 31–36. doi: 10.1007/978-981-10-6747-1_4.

[17] O. Maimon and L. Rokach, *Soft computing for knowledge discovery and data mining*. Springer Science &

Business Media, 2007. doi: 10.1007/978-0-387-69935-6.
[18] E. Pashaei, M. Ozen, and N. Aydin, "Improving medical diagnosis reliability using Boosted C5. 0 decision tree empowered by Particle Swarm Optimization," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 7230–7233. doi: 10.1109/EMBC.2015.7320060.
[19] Z. Mohammadi-Pirouz, K. Hajian-Tilaki, M. Sadeghi Haddat-Zavareh, A. Amoozadeh, and S. Bahrami, "Development of decision tree classification algorithms in predicting mortality of COVID-19 patients," *Int. J. Emerg. Med.*, vol. 17, no. 1, p. 126, 2024, doi: 10.1186/s12245-024-00681-7.
[20] Z. Azam, M. M. Islam, and M. N. Huda, "Comparative analysis of intrusion detection systems and machine learning-based model analysis through decision tree," *Ieee Access*, vol. 11, pp. 80348–80391, 2023, doi: 10.1109/ACCESS.2023.3296444.
[21] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *J. Appl. Sci. Technol. trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
[22] H. Oh and S. Lee, "Evaluation and interpretation of tourist satisfaction for local Korean festivals using explainable AI," *Sustainability*, vol. 13, no. 19, p. 10901, 2021, doi: 10.3390/su131910901.
[23] M. J. Sánchez-Franco and S. Rey-Tienda, "The role of user-generated content in tourism decision-making: an exemplary study of Andalusia, Spain," *Manag. Decis.*, vol. 62, no. 7, pp. 2292–2328, 2024, doi: 10.1108/MD-06-2023-0966.
[24] A. García-Milon, E. Juaneda-Ayensa, C. Olarte-Pascual, and J. Pelegrín-Borondo, "Towards the smart tourism destination: Key factors in information source use on the tourist shopping journey," *Tour. Manag. Perspect.*, vol. 36, p. 100730, 2020, doi: 10.1016/j.tmp.2020.100730.
[25] P. Liberato, E. Alen, and D. Liberato, "Smart tourism destination triggers consumer experience: the case of Porto," *Eur. J. Manag. Bus. Econ.*, vol. 27, no. 1, pp. 6–25, 2018, doi: 10.1108/EJMBE-11-2017-0051.
[26] C. Wang, H. Qu, and M. K. Hsu, "Toward an integrated model of tourist expectation formation and gender difference," *Tour. Manag.*, vol. 54, no. 6, pp. 58–71, 2016, doi: 10.1016/j.tourman.2015.10.009.
[27] S. Khademi, F. Rafieinasab, and N. Radhakrishnan, "Data mining in academic libraries: a systematic review," *Int. J. Inf. Sci. Manag.*, vol. 21, no. 3, pp. 255–271, 2023.
[28] K. Kowsari, K. J. Meimandi, M. Heidarysafa, and S. Mendu, "Text Classification Algorithms : A Survey," *Information*, vol. 10, no. 4, pp. 1–68, 2019, doi: 10.3390/info10040150.
[29] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *ieee Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, 2018, doi: 10.1109/MCI.2018.2840738.
[30] S. M. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017, no. Section 2, pp. 1–10. doi: 10.48550/arXiv.1705.07874.
[31] M. T. Ribeiro, "' Why Should I Trust You ?' Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2026, no. March, pp. 1135–1144. doi: 10.1145/2939672.2939778.
[32] R. Guidotti, "A Survey of Methods for Explaining Black Box Models A Survey of Methods for Explaining Black Box Models," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–93, 2026, doi: 10.1145/3236009.
[33] Y. Xia *et al.*, "Sustainable service quality assessment of Chinese healthcare e-government: a multi-criteria decision framework based on SERVQUAL model and entropy-weight TOPSIS method," *Front. Digit. Heal.*, vol. 7, p. 1611979, 2025, doi: 10.3389/fdgth.2025.1611979.