Volume XI Issue 2 Year 2026 | Page 527-538 | ISSN: 2527-9866
Received: 26-02-206 | *Revised: 04-04-2026* | *Accepted: 26-05-2026*

Sentiment Analysis of Tourists' Perceptions of Ubud as a World Gastronomy Destination Using the Lexicon-SVM Method

Ni Wayan Sumartini Saraswati[1], I Wayan Dharma Suryawan[2], I Kadek Agus Bisena[3], I Dewa Made Krishna Muku[4], Dewa Ayu Kadek Pramita[5]
[1,2,3,4,5]Institut Bisnis dan Teknologi Indonesia, Panjer, Denpasar, Indonesia, 80225
e-mail: sumartini.saraswati@gmail.com[1], dharma.suryawan@gmail.com[2], agus.bisena@gmail.com[3], krishna.muku@gmail.com[3], pramita.wayu@gmail.com[3]
*Correspondence: *sumartini.saraswati@gmail.com*

Abstract: *The development of Ubud as a sustainable gastronomic tourism destination requires understanding tourist perceptions expressed on digital platforms. This study analyzes tourist sentiment toward Ubud's gastronomy using English-language reviews from TripAdvisor and X (formerly Twitter) through a hybrid Lexicon–Support Vector Machine (SVM) approach. A total of 28,550 textual data were analyzed, consisting of 23,647 TripAdvisor reviews and 4,903 X posts. The methodology includes data collection, text preprocessing, sentiment labeling using the VADER lexicon, TF-IDF feature extraction, and SVM classification. Model performance was evaluated using accuracy, precision, recall, and F1-score. The results show that positive sentiment dominates on both platforms, with accuracies of 89.7% for X and 92.31% for TripAdvisor. Word cloud analysis further indicates that tourist perceptions are influenced by food quality, service, atmosphere, and pricing. These findings demonstrate the potential of the hybrid Lexicon–SVM approach for supporting sustainable gastronomic tourism development in Ubud. The study also contributes comparative insights into sentiment characteristics between structured reviews and real-time social media platforms.*

Keywords: *Ubud, Gastronomic, Sentiment Analysis, Lexicon, Support Vector Machine.*

1. Introduction

Ubud is a tourism destination with a strong international reputation, especially because of its rich authentic arts and culture, natural beauty, and the friendliness of the local people. Along with ongoing efforts to preserve cultural heritage and implement sustainable tourism practices, Ubud has long been known as a leading destination for domestic and international tourists [1]. To further strengthen its competitiveness, the Ministry of Tourism and Creative Economy (Kemenparekraf), in collaboration with the Gianyar Regional Government and the United Nations World Tourism Organization (UNWTO), initiated the development of Ubud as a gastronomic tourism destination [2]. Ubud was chosen as a pilot project because of its infrastructure readiness, strong culinary traditions, and collaborative stakeholder ecosystem.

As part of this initiative, the UNWTO completed a gastronomic tourism development project aimed at positioning Ubud as a global center for sustainable gastronomic tourism [3]. Gastronomy is not only seen as a tourism product but also as part of cultural heritage that can encourage local economic growth, job creation, and inclusive development. Through gastronomic experiences, tourists not only interact with culinary products but also with cultural values and local identity contained in the process of serving and processing food. In this way, Ubud further strengthens its position on the global tourism map as a culinary destination and a center for gastronomic education rooted in local culture [4].
Tourists' perceptions of Ubud as a gastronomic destination are increasingly documented on various digital platforms such as TripAdvisor and X (formerly Twitter). TripAdvisor provides structured and relatively comprehensive reviews of food quality, service, and atmosphere,

making it a valuable source for understanding tourist satisfaction and preferences [5]. Meanwhile, X enables spontaneous and real-time expression of opinions, reflecting direct responses to gastronomic experiences [6]. The information generated from these platforms is a significant source of data for the tourism and culinary industries because it can reveal insights regarding tourist perceptions, service quality, and emerging culinary trends [7].

However, despite the increasing abundance of digital data related to tourist perceptions, most previous research is still dominated by descriptive analysis approaches or the use of a single sentiment analysis method, either lexicon-based or machine learning-based, each of which has limitations in capturing the context and complexity of meaning in textual data [8]. Furthermore, studies specifically examining tourist perceptions of gastronomic tourism still tend to focus on a single data source, thus failing to fully represent the dynamics of tourist opinion, especially when comparing the characteristics of structured data (such as TripAdvisor) with unstructured, real-time data (such as X) [9]. In the context of a culture-based destination such as Ubud, this limitation becomes even more significant because tourist perceptions are influenced not only by culinary quality but also by the cultural dimensions inherent in the gastronomic experience.

A significant research gap exists regarding the limitations of sentiment analysis approaches in accurately accommodating the complexity of textual data, as well as the lack of multi-platform data integration in analyzing tourist perceptions of culture-based gastronomic destinations. This research presents novelty through the application of a hybrid lexicon–SVM approach that combines the advantages of lexicon-based methods in capturing semantic polarity with the ability of Support Vector Machines (SVM) to perform pattern-based classification more precisely [9], thereby overcoming the limitations of each method when used separately. The integration of data from two digital platforms with different characteristics, namely TripAdvisor and X, allows for a more comprehensive and contextual analysis of tourist perceptions. The contribution of this research includes methodological aspects in the development of a hybrid-based sentiment analysis approach as well as empirical aspects in enriching gastronomic tourism studies, especially in understanding tourist perceptions of culture-based destinations such as Ubud in a more holistic and data-driven way.

2. Literature Review

Sentiment analysis has developed into a widely adopted approach for extracting insights from user-generated textual data, particularly in tourism research. In general, sentiment analysis is conducted using two primary approaches: machine learning-based methods and lexicon-based methods [10]. Machine learning algorithms, especially Support Vector Machines (SVM), have demonstrated strong capability in handling high-dimensional text data. Previous studies report that SVM performs effectively in classifying sentiments within tourism-related contexts, including hotels, restaurants, and tourist destinations [11], [12], [13]. Furthermore, SVM has been shown to remain competitive even when compared to deep learning approaches, which typically require large datasets and substantial computational resources [14]. These findings highlight the robustness and efficiency of SVM for sentiment classification tasks involving structured and unstructured tourism data.

On the other hand, lexicon-based approaches rely on predefined sentiment dictionaries to determine the polarity of textual data. Methods such as the VADER Lexicon are widely used because they do not require labeled training data and are relatively straightforward to implement. This makes lexicon-based approaches particularly attractive for large-scale data analysis. However, several studies indicate that lexicon-only approaches may face limitations in capturing domain-specific terminology, contextual nuances, and informal language variations, which can lead to lower classification accuracy compared to supervised machine learning models [15]. To address these limitations, recent research has explored hybrid models that combine lexicon-based techniques with machine learning algorithms. Studies integrating lexicon methods with SVM demonstrate improved efficiency in data labeling while maintaining strong classification performance [16], [17], [18]. Hybrid approaches leverage the automatic polarity scoring capability of lexicon methods and the pattern-learning strength of SVM, thereby enhancing model reliability and scalability in large textual datasets.

Despite the increasing application of SVM-based and hybrid sentiment analysis in tourism research, most prior studies focus primarily on general tourism services such as hotels and tourist destinations. Limited attention has been given to gastronomic tourism as a representation of a destination's cultural identity. Consequently, a gap remains in applying hybrid lexicon–SVM approaches specifically to analyze tourist perceptions of gastronomic tourism contexts. Addressing this gap is essential to strengthen methodological development in tourism analytics while deepening understanding of gastronomy as a cultural and experiential dimension of tourism.

3. Methods

The research flow in this study is shown in Figure 1, which systematically illustrates the research stages from data collection to model evaluation. The initial stage began with data collection from two major digital platforms, TripAdvisor and X (formerly Twitter), which represent different data characteristics: structured reviews and unstructured real-time opinions. The dataset used consisted of 23,647 TripAdvisor reviews and 4,903 tweets relevant to tourists' perceptions of Ubud as a gastronomic destination. Data collection was conducted online using web scraping techniques and APIs, where Selenium was used to extract data from TripAdvisor, and the Twitter API was utilized to obtain data from X based on keywords related to Ubud gastronomy, as shown in Table 1. The data timeframe differs for each platform, namely September 2011 to September 2021 for TripAdvisor and June 2018 to December 2024 for X, due to limited data access on TripAdvisor. To ensure data quality and consistency, a filtering and cleaning process was carried out by removing duplicate data, retaining only English text, and filtering out irrelevant and noisy content, such as uninformative text or symbols, resulting in a more representative dataset. The combination of these two data sources provides comprehensive coverage, with TripAdvisor representing in-depth, experience-based reviews and X reflecting spontaneous opinions and current trends.

Table 1. Examples of Query Keywords Used for Data Collection from X

No.	Keyword	No.	Keyword	No.	Keyword
1.	Food lovers Ubud	7.	Ubud Food Rating	13.	Ubud Food
2.	Travel Ubud	8.	Ubud Food Recommendation	14	Ubud Food Aesthetic
3.	Ubud Food Tourism	9.	Ubud Food Review	15.	Ubud Food Culture
4.	Ubud Authentic Food	10.	Ubud Food Spots	16.	Ubud Food Destination
5.	Ubud Best Food	11.	Ubud Food Tour	17.	Ubud Food Experience
6.	Ubud Culinary Art	12.	Ubud Food Tradition	18.	Ubud Local Food

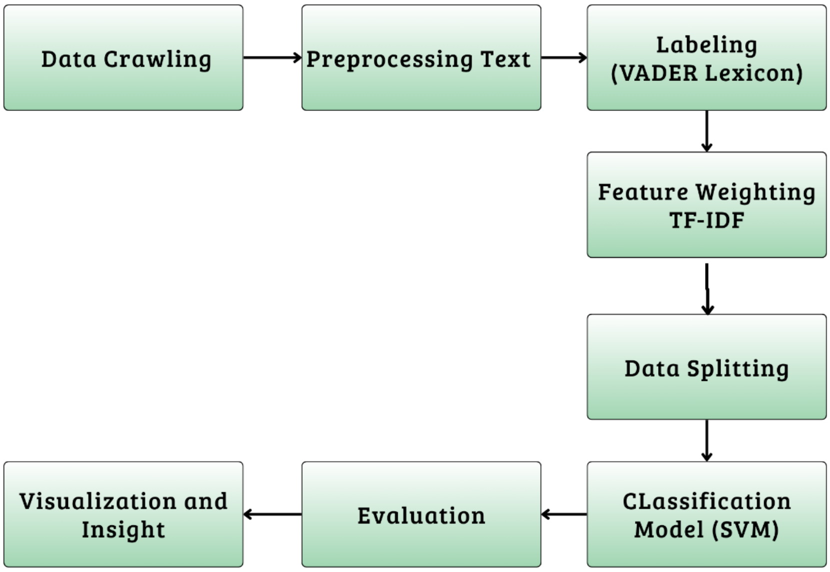

Figure 1. Research Flow

A. Data Preprocessing

The collected text data then undergo a preprocessing stage to improve data quality before further analysis. This process was carried out in stages, including cleaning, case folding, tokenization, stopword removal, and lemmatization. In the cleaning stage, various irrelevant elements are removed, such as URLs, mentions (@user), hashtags, emojis, numbers, punctuation, and non-alphabetic characters commonly found in social media data, so that the text becomes cleaner and more structured [19]. Next, case folding was performed by converting all text to lowercase to standardize the writing format [20]. The tokenization stage breaks the text into word units to facilitate analysis [21], and then continues with stopword removal to eliminate common words that do not have a significant contribution to sentiment analysis, such as "the" and "and" [22]. In the final stage, lemmatization was performed to convert words to their basic form so that word variations could be represented consistently. The entire preprocessing process was performed using the Natural Language Toolkit (NLTK) library in the Python programming language.

B. Lexicon-Based Labeling

After preprocessing, the text data were sentiment-labeled using a lexicon-based approach with the Valence Aware Dictionary and Sentiment Reasoner (VADER) method, which is designed for analyzing informal texts such as reviews and social media. VADER generates a polarity score based on compound values, which are used to classify sentiment into two categories: positive and negative. In this study, text with a compound score ≥ 0.05 was categorized as positive, while a score ≤ -0.05 was categorized as negative, and text with a neutral score was simplified into the positive category to support a binary classification scheme [23]. This approach is used as a form of pseudo-labeling to automatically generate initial labels on large datasets, making it more efficient than manual labeling alone. However, this approach has limitations in capturing context and certain linguistic nuances, potentially leading to error propagation in the supervised model. To improve reliability, particularly for TripAdvisor data, a user-rating-based ground truth was also used, where ratings of 4–5 were categorized as positive and 1–2 as negative, allowing validation of the lexicon-based labeling results. Thus, this combination of approaches provides a balance between the labeling efficiency and quality of sentiment representation in the dataset.

C. TF-IDF Feature Extraction

The sentiment-labeled text data were then transformed into numerical form using the TF-IDF method. This method is used to measure the importance of a word in a document relative to the entire corpus [24]. Mathematically, the TF-IDF weighting is formulated as follows (1).

$$TF - IDF(t, d) = TF(t, d) \times log\frac{N}{DF(t)} \quad (1)$$

where $TF(t, d)$ is the frequency of word t in document d, N is the total number of documents, and $DF(t)$ is the number of documents that contain word t.

D. Support Vector Machine Classification

Before the classification process was carried out, the TF-IDF weighted data were divided into training and test data with a ratio of 80:20. Sentiment classification was performed using the SVM algorithm, which aims to construct an optimal hyperplane to separate positive and negative sentiment classes with a maximum margin [25], as illustrated in Figure 2. Each data point represents a review that has been transformed into a feature space, whereas the points closest to the decision boundary, called support vectors, play a crucial role in determining the position of the hyperplane. By maximizing the margin between classes, SVM can improve the model's generalization ability in classifying sentiments in high-dimensional text data, such as TF-IDF extraction results [26], [27].

Figure 2 shows the basic concept of SVM in forming an optimal hyperplane to separate two sentiment classes. The center line represents the decision boundary, and the two parallel lines indicate the margin formed between the positive and negative classes. Data points closest to the decision boundary are called support vectors and play a crucial role in determining the position of the hyperplane. SVM works by maximizing the margin between classes, thereby improving the model's generalization ability in classifying high-dimensional text data [28]. This method is also known to be effective in text classification tasks because it can optimally handle high-dimensional features and produce competitive classification performance [29].

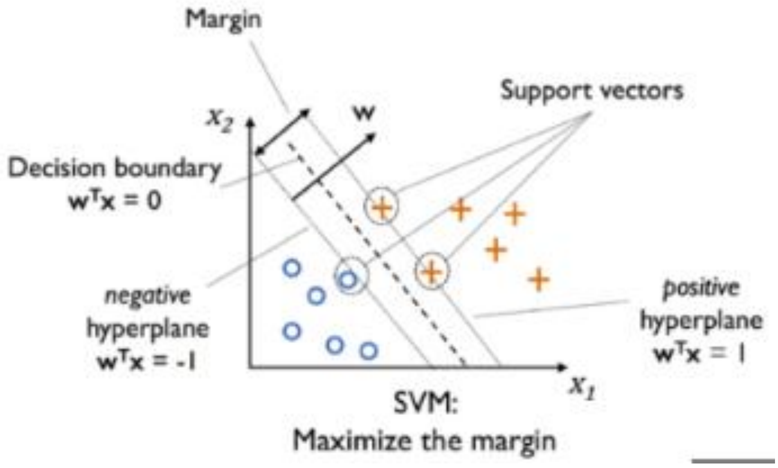

Figure 2. Support Vector Machine

E. Model Evaluation

The model performance was evaluated using a confusion matrix. The confusion matrix consists of four components: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) [30]. Based on the confusion matrix, several evaluation metrics were calculated, including accuracy, precision, recall, and f1-score, formulated as follows (2)–(5).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

$$F1\text{-}score = 2 \times \frac{Precision \times recall}{Precision + recall} \times 100\% \quad (5)$$

F. Wordcloud Visualization

As the final stage, the sentiment analysis results were visualized in the form of a word cloud, which displays the frequency of word occurrences in the data [31]. The more frequently a word appears, the larger its size in visualization. A word cloud was used to identify dominant words with positive and negative sentiments, thus providing insight into aspects of gastronomy that need to be maintained and those that require evaluation as a basis for formulating strategies for developing Ubud's gastronomic tourism.

4. Results and Discussion

A. Distribution of Tourist Sentiment

The dataset used in this study consisted of English-language tourist reviews sourced from the TripAdvisor platform and the social media platform X (Twitter). After pre-processing and sentiment labeling using the VADER Lexicon method, the data were classified into two main categories: positive and negative sentiments. The distribution of the sentiment labeling results from both data sources is presented in Table 1. As shown in Table 1, positive sentiment dominated tourist reviews on both platforms. TripAdvisor shows a positive sentiment proportion of 96.07%, whereas the social media platform X has a positive sentiment proportion of 95.42%. This dominance of positive sentiment indicates that the gastronomic experience in Ubud generally receives a positive response from tourists. However, negative sentiments were still identified in both data sources, reflecting certain aspects that still require further attention and evaluation in the development of gastronomic tourism.

Table 2. Distribution of Sentiment Results

Data Sources	Positive Sentiment	Positive (%)	Negative Sentiment	Negative (%)	Total Data
X (Twitter)	4,678	95.42%	225	4.58%	4,903
TripAdvisor	22,719	96.07%	928	3.93%	23,647
Total	27,397	95.96%	1,153	4.04%	28,550

B. SVM Classification Model Evaluation

The model performance was evaluated using a confusion matrix and standard evaluation metrics: accuracy, precision, recall, and F1-score. A summary of the evaluation results for each data source is presented in Table 2, while more detailed classification results are shown in the confusion matrices in Figures 3 (dataset X) and 4 (TripAdvisor dataset). According to Table 2, the SVM model achieved an accuracy of 89.70% on dataset X and 92.31% on TripAdvisor dataset X. The resulting recall values are very high, reaching 100% on dataset X and 99.94% on the TripAdvisor dataset, indicating that almost all data with positive sentiments were correctly classified. However, the analysis of the confusion matrices in Figures 3 and 4 shows that these high recall values are accompanied by a relatively large number of false positives, where data with negative sentiments are classified as positive. This pattern indicates that the model tends to be biased toward the majority class, namely, positive sentiment, which dominates both datasets.

Table3. SVM Classification Model Evaluation Results

Sumber Data	Accuracy	Precission	Recall	F1-score
X	89.70%	89.68%	100%	94.56%
TripAdvisor	92.31%	92.33%	99.94%	95.99%

In dataset X (Figure 3), all positive data were correctly classified (FN = 0); however, there were still misclassifications in the negative class (FP = 42), indicating the model's limitations in distinguishing a smaller number of negative sentiments. A similar pattern was also observed in the TripAdvisor dataset (Figure 4), where the number of correctly classified positive data was very high (TP = 17,824), but the misclassification of negative data was still quite significant (FP = 1,480). This indicates that class imbalance affects the model performance, particularly in identifying sentiments in minority classes.

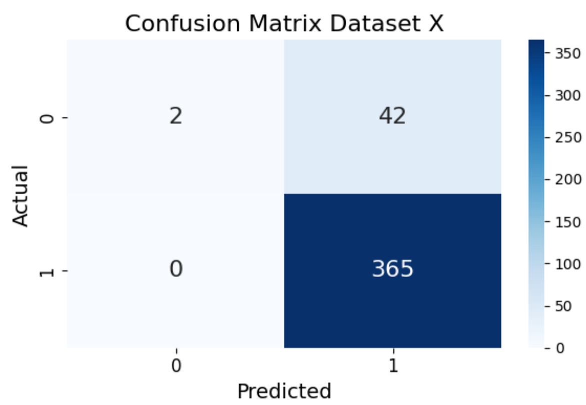

Figure 3. Confusion Matrix of SVM Classification Results on X Dataset

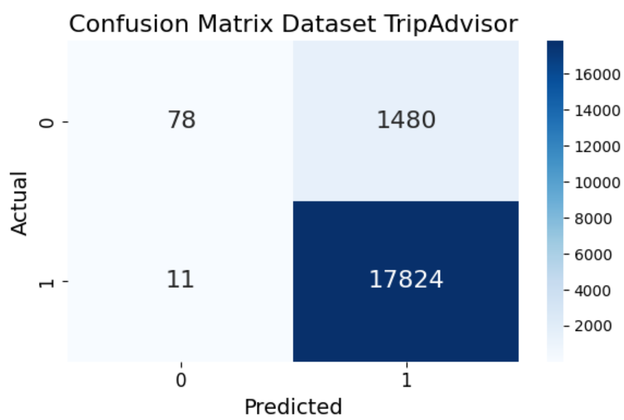

Figure 4. Confusion Matrix of SVM Classification Results on TripAdvisor Dataset

Furthermore, the higher performance of the TripAdvisor dataset compared to X is influenced not only by the more structured and descriptive characteristics of the data but also by more consistent language patterns in user reviews. In contrast, data from X tend to be shorter, more informal, and more varied, increasing the noise and complexity of feature representation. Thus, although the SVM model demonstrates a good ability to capture dominant sentiment patterns in high-dimensional text data, the results obtained need to be interpreted considering data characteristics and class imbalance, rather than solely based on aggregate accuracy and recall values.

C. Wordcloud Visualization Analysis

According to Figure 3, the word cloud visualization of social media data X shows that positive sentiment is characterized by the dominance of phrases such as “the best,” “food in Ubud,” “guide to,” and operational-related expressions such as “open daily” and “from morning to night.” These patterns indicate that positive conversations are largely centered on culinary recommendations, perceived quality of food, and accessibility of dining options. The presence of informative expressions suggests that users not only share satisfaction but also provide references and guidance for other tourists.

Conversely, negative sentiment in social media data X is dominated by phrases such as “dangerous food,” “extreme food,” “extreme Balinese food,” and terms related to raw ingredients such as “raw meat” and “blood.” This indicates that negative perceptions are primarily associated with concerns regarding food safety and the perceived extremity of certain traditional culinary experiences. This finding is supported by representative user expressions such as “the food was interesting but too extreme for me,” which reflect that negative sentiment is not necessarily driven by taste, but by tourists’ unfamiliarity with local culinary characteristics.

Overall, the contrast between positive and negative word patterns in dataset X reflects differing thematic emphases, where positive sentiment relates to food quality and accessibility, while negative sentiment is associated with perceived risk and culinary unfamiliarity.

Figure 5. Word Cloud of Platform X: (a) Positive Bigrams, (b) Negative Bigrams, (c) Positive Trigrams, (d) Negative Trigrams.

Based on Figure 4, the word cloud visualization of TripAdvisor review data shows that positive sentiment is dominated by phrases related to food and service quality, such as “good food,” “great food,” “friendly staff,” “good service,” and “highly recommended.” Additionally, expressions such as “good value for money,” “rice field view,” and “monkey forest” indicate that tourist satisfaction is influenced not only by food quality but also by environmental and experiential aspects. This suggests that TripAdvisor reviews tend to be more evaluative and structured, emphasizing the overall dining experience.

In contrast, negative sentiment in TripAdvisor data is dominated by phrases related to service performance and time efficiency, such as “bad food experience,” “bad service,” “wait minutes,” and “waste time.” The presence of phrases such as “pay bill,” “leave,” and “return” indicates dissatisfaction with service processes and post-experience decisions. This pattern is further supported by representative reviews such as “the food was good but the service was very slow,” highlighting that negative sentiment often arises from a mismatch between service quality and expectations rather than food quality alone.

From a thematic perspective, the word cloud analysis across both platforms can be categorized into several key dimensions, namely food quality, service quality, dining experience, and food safety perception. These dimensions collectively shape tourist perceptions of gastronomic experiences in Ubud. Compared to social media X, TripAdvisor reviews provide more consistent and descriptive linguistic patterns, enabling clearer identification of evaluation aspects, while X reflects more spontaneous and diverse expressions that capture immediate reactions and emerging perceptions.

Figure 6. Word Cloud of Platform TripAdvisor: (a) Positive Bigrams, (b) Negative Bigrams, (c) Positive Trigrams, (d) Negative Trigrams.

The results indicate that the SVM-based sentiment classification model, combined with lexicon-based pre-labeling, demonstrates strong performance on both data sources, as reflected in relatively high accuracy, precision, recall, and F1-score values. However, as shown in the confusion matrix analysis, this performance is influenced by the dominance of positive sentiment in the dataset, which leads to a tendency of the model to favor the majority class. Consequently, while the model is effective in identifying dominant sentiment patterns, its performance in capturing minority (negative) sentiments remains limited. The higher performance observed in the TripAdvisor dataset compared to social media platform X suggests that data characteristics play an important role in classification outcomes. The more structured and descriptive nature of TripAdvisor reviews enables more consistent pattern learning, whereas the shorter, more informal, and linguistically diverse nature of social media data increases classification complexity.

The word cloud analysis further complements these findings by revealing thematic differences in sentiment expression across platforms. Positive sentiment is primarily associated with food quality, service, and overall dining experience, while negative sentiment is more closely related to service issues, waiting time, and perceptions of food safety or unfamiliar culinary experiences. These findings highlight that tourist perceptions of gastronomic experiences are shaped by multiple dimensions, not limited to food quality alone. Overall, this study suggests that sentiment analysis can provide valuable insights into tourist perceptions and can support data-driven evaluation in the development of gastronomic tourism in Ubud, while still requiring careful interpretation in relation to data characteristics and class distribution.

5. Conclusions

This study demonstrates that a hybrid approach combining lexicon-based sentiment labeling using VADER and SVM classification can capture sentiment patterns in tourists' perceptions of Ubud's gastronomy. The SVM model, applied to TF-IDF feature representation, performed well in classifying positive and negative sentiments in data from social media platforms X and TripAdvisor, with relatively better performance on more structured data. The results of sentiment analysis and word cloud visualization indicate that tourists' perceptions of Ubud's gastronomy are dominated by positive sentiment, particularly regarding food quality, service, and the overall dining experience. Meanwhile, negative sentiment generally relates to service aspects, wait times, and perceptions of culinary experiences as unusual or unfamiliar, indicating areas that still need improvement to enhance tourist satisfaction. It should be noted that the model performance is affected by an imbalanced class distribution, where positive sentiment dominates the dataset. This condition tends to favor the majority class, thus limiting its ability to identify negative sentiments as a minority class. Therefore, the interpretation of the results must consider these data characteristics. This research provides insights that stakeholders can leverage to formulate data-driven strategies for developing gastronomic tourism in Ubud, particularly in improving service quality and managing tourist expectations regarding local culinary characteristics. Furthermore, this study contributes to the development of sentiment analysis applications in tourism, particularly through the integration of lexicon-based approaches and machine learning.

References

[1] N. M. P. Bestari, A. Suryawan Wiranatha, I. G. A. Oka Suryawardani, and I. N. Darma Putra, "Rejuvenating Cultural Tourism Through Gastronomic Creative Tourism in Ubud Bali," *Mudra Jurnal Seni Budaya*, vol. 37, no. 2, pp. 136–145, Apr. 2022, doi: 10.31091/mudra.v37i2.1938.

[2] kemenparekraf, "Siaran Pers : Menparekraf: Ubud Akan Jadi Pusat Prototipe Gastronomi Dunia," 2023. Accessed: Jan. 29, 2026. [Online]. Available: https://www.kemenparekraf.go.id/berita/siaran-pers-menparekraf-ubud-akan-jadi-pusat-prototipe-gastronomi-dunia

[3] UNWTO, "UNWTO Completes Gastronomy Tourism Project for Ubud, Indonesia." Accessed: Jan. 29, 2026. [Online]. Available: https://www.untourism.int/news/unwto-completes-gastronomy-tourism-project-for-ubud-indonesia

[4] N. N. S. Aryanti, E. Harmayani, W. Udasmoro, and D. Mutiarin, "Habitus and Capital in the Gastronomy Tourism Field: Cooking Class Businesses in Ubud, Bali," *E-Journal of Tourism*, vol. 12, no. 2 SE-Articles, pp. 285–297, Sep. 2025, doi: 10.24922/eot.v12i2.2184.

[5] TripAdvisor, "Menurut wisatawan, tempat terbaik untuk menemukan ulasan berkualitas adalah Tripadvisor." Accessed: Jan. 31, 2026. [Online]. Available: https://www.tripadvisor.com/business/id/insights/content-api/resources/travelers-say-the-the-best-place-to-find-quality-reviews-is-on-tripadvisor?

[6] A. H. Al Fajar and Dyas Erlangga, "The Role of Social Media and User-Generated Content in Traveler Decisions: A Literature Review," *Journal of Tourism and Creativity*, vol. 9, no. 1, pp.

91–98, May 2025, doi: 10.19184/jtc.v9i1.53690.

[7] K. M. Aboalganam, S. F. AlFraihat, and S. Tarabieh, “The Impact of User-Generated Content on Tourist Visit Intentions: The Mediating Role of Destination Imagery,” *Administrative Sciences*, vol. 15, no. 4, p. 117, Mar. 2025, doi: 10.3390/admsci15040117.

[8] P. Prihananto, L. R. Noer, S. Y. Ninglasari, and N. G. M. Rai, “Satisfaction and User Reviews in Tourism Using Big Data and Text Mining: A Bibliometric Study,” *International Journal of Academic Research in Business and Social Sciences*, vol. 14, no. 12, pp. 2194–2210, Dec. 2024, doi: 10.6007/IJARBSS/v14-i12/24181.

[9] K. L. Tan, C. P. Lee, and K. M. Lim, “A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research,” *Applied Sciences*, vol. 13, no. 7, p. 4550, Apr. 2023, doi: 10.3390/app13074550.

[10] Y. Qi and Z. Shabrina, “Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach,” *Social Network Analysis and Mining*, vol. 13, no. 1, p. 31, Feb. 2023, doi: 10.1007/s13278-023-01030-x.

[11] M. Tabany and M. Gueffal, “Sentiment Analysis and Fake Amazon Reviews Classification Using SVM Supervised Machine Learning Model,” *Journal of Advances in Information Technology*, vol. 15, no. 1, pp. 49–58, 2024, doi: 10.12720/jait.15.1.49-58.

[12] N. Pavitha *et al.*, “Movie recommendation and sentiment analysis using machine learning,” *Global Transitions Proceedings*, vol. 3, no. 1, pp. 279–284, Jun. 2022, doi: 10.1016/j.gltp.2022.03.012.

[13] N. Leelawat *et al.*, “Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning,” *Heliyon*, vol. 8, no. 10, p. e10894, Oct. 2022, doi: 10.1016/j.heliyon.2022.e10894.

[14] K. Puh and M. Bagić Babac, “Predicting sentiment and rating of tourist reviews using machine learning,” *Journal of Hospitality and Tourism Insights*, vol. 6, no. 3, pp. 1188–1204, Jun. 2023, doi: 10.1108/JHTI-02-2022-0078.

[15] K. Barik and S. Misra, “Analysis of customer reviews with an improved VADER lexicon classifier,” *Journal of Big Data*, vol. 11, no. 1, p. 10, Jan. 2024, doi: 10.1186/s40537-023-00861-x.

[16] R. Obiedat *et al.*, “Sentiment Analysis of Customers’ Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution,” *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.

[17] R. H. Muhammadi, T. G. Laksana, and A. B. Arifa, “Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis,” *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 8, no. 1, pp. 59–71, Mar. 2022, doi: 10.23917/khif.v8i1.15213.

[18] N. Saraswathi, T. Sasi Rooba, and S. Chakaravarthi, “Improving the accuracy of sentiment analysis using a linguistic rule-based feature selection method in tourism reviews,” *Measurement: Sensors*, vol. 29, p. 100888, Oct. 2023, doi: 10.1016/j.measen.2023.100888.

[19] M. A. Lones, “Avoiding common machine learning pitfalls,” *Patterns*, vol. 5, no. 10, p. 101046, Oct. 2024, doi: 10.1016/j.patter.2024.101046.

[20] N. Cahyono and Anggista Oktavia Praneswara, “Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes,” *Indonesian Journal of Computer Science*, vol. 12, no. 6, Dec. 2023, doi: 10.33022/ijcs.v12i6.3473.

[21] B. S. Gunawan, E. Evanita, and A. A. Riadi, “Analisis Sentimen Terhadap Ulasan Aplikasi Bsi Mobile Menggunakan Algoritma Naïve Bayes Classifier,” *Bina Informatika dan Komputer (BINER)*, vol. 2, no. 1, May 2024, doi: 10.24176/biner.v2i1.13399.

[22] M. Akbar and W. Yustanti, “Pemilihan Algoritma Klasifikasi Terbaik Untuk Prediksi Jenis Keluhan MI User Interface (MIUI) 14,” *Journal of Informatics and Computer Science (JINACS)*, vol. 6, no. 02, pp. 445–452, 2024.

[23] Y. A. Singgalen, “Sentiment and Toxicity Analysis of Digital Content Using Perspective, Vader, and TextBlob: Tourism and Birdwatching,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 5, no. 1, pp. 142–153, 2024, doi: 10.30865/klik.v5i1.2091.

[24] N. Fibriyanti Arminda, N. Sulistiyowati, and T. Nur Padilah, “Implementasi Algoritma Multinomial Naive Bayes Pada Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Brimo,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 3, pp. 1817–1822, Nov. 2023, doi: 10.36040/jati.v7i3.7012.

[25] S. Ernawati and R. Wati, "Evaluasi Performa Kernel SVM dalam Analisis Sentimen Review Aplikasi ChatGPT Menggunakan Hyperparameter dan VADER Lexicon," *Jurnal Buana Informatika*, vol. 15, no. 01, pp. 40–49, Apr. 2024, doi: 10.24002/jbi.v15i1.7925.

[26] M. R. Fahlevvi, "Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi Dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia Di Google Playstore Menggunakan Metode Support Vector Machine," *Jurnal Teknologi dan Komunikasi Pemerintahan*, vol. 4, no. 1, pp. 1–13, Jun. 2022, doi: 10.33701/jtkp.v4i1.2701.

[27] R. Pratiwi, D. Dairoh, D. A.-J. of Informatics, and U. 2021, "Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM)," *journal.ittelkom-pwt.ac.idRW Pratiwi, D Dairoh, DI Af'idahJournal of Informatics Information System Software, 2021•journal.ittelkom-pwt.ac.id*, 2021, doi: 10.20895/INISTA.V4I1.

[28] T. Jung and J. Kim, "A new support vector machine for categorical features," *Expert Systems with Applications*, vol. 229, p. 120449, 2023, doi: 10.1016/j.eswa.2023.120449.

[29] Q. Wang *et al.*, "Deep convolutional cross-connected kernel mapping support vector machine based on SelectDropout," *Information Sciences*, vol. 626, pp. 694–709, 2023, doi: 10.1016/j.ins.2023.01.004.

[30] G. Zeng, "Invariance Properties and Evaluation Metrics Derived from the Confusion Matrix in Multiclass Classification," *Mathematics*, vol. 13, no. 16, p. 2609, Aug. 2025, doi: 10.3390/math13162609.

[31] M. Dong, J. Lu, G. Wang, X. Zheng, and D. Kiritsis, "Model-based Systems Engineering Papers Analysis based on Word Cloud Visualization," in *2022 IEEE International Systems Conference (SysCon)*, IEEE, Apr. 2022, pp. 1–7. doi: 10.1109/SysCon53536.2022.9773795.

Acknowledgements

We thank INSTIKI Research and Development Program in 2026 for the support, both morally and financially, for the successful implementation of this research.