E-Jurnal Matematika Vol. 15(2), Mei 2026, pp. 48-58
DOI: https://doi.org/10.24843/MTK.2026.v15.i02.p502
ISSN: 2303-1751

COMPARATIVE EVALUATION OF SVM AND LSTM FOR TOURISM SENTIMENT CLASSIFICATION: STUDY CASE TANAH LOT TOURISM DESTINATION, BALI.

Dewa Made Alit Adinugraha[1§], Jery Christianto[2]

[1]Politeknik Internasional Bali [Email:alit.adinugraha@pib.ac.id]
[2]Politeknik Internasional Bali [Email: jery.christianto@pib.ac.id]
[§]*Corresponding Author*

ABSTRACT

This study presents a comparative evaluation of Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) models for tourism sentiment classification, using YouTube comments related to Tanah Lot, Bali. The dataset manually cleaned comments labeled as Positive, Neutral, or Negative. Both models achieved identical overall accuracy (0.95), but class-wise analysis revealed substantial differences: LSTM exhibited strong bias toward the majority class (Neutral), failing to detect minority sentiments, while SVM retained partial sensitivity to Positive and Negative classes. These findings highlight the limitations of deep learning architectures under low-resource and imbalanced conditions and underscore the importance of context-aware model selection. Class-wise evaluation metrics are essential for capturing minority sentiment, which is critical for destination governance and informed decision-making in tourism management.

Keywords: *Class Imbalance, Destination Governance, Tanah Lot, User-Generated Content (UGC)*

1. INTRODUCTION

Digital transformation has shifted the structure of the tourism industry from a conventional information distribution model to a data-driven ecosystem and user participation. The evolution of information technology in the last two decades has not only improved the efficiency of reservation and marketing systems but also changed the architecture of interaction between tourists and destinations (Buhalis & Law, 2008; Law et al., 2014). The concept of smart tourism emphasizes the integration of data, technology, and analytics as the foundation of modern destination governance (Gretzel et al., 2015; Wang et al., 2013).

This transformation expands the role of tourists as active actors in the production of destination value through *a digital-based co-creation* process (Dang & Nguyen, 2023). Digitalization is no longer just a promotional instrument, but a mechanism for forming a collective perception of a destination (Kindzule-Millere & Zeverte-Rivza, 2022). Thus, digital data generated by tourists becomes a structural component in destination management.

User-generated content (UGC) has become the primary source of information in travel decision-making. Electronic word-of-mouth (e-WOM) has been shown to have a significant influence on restaurant popularity, hotel performance, and destination image (Litvin et al., 2008; Zhang et al., 2010; Kim et al., 2016). Travel blogs, online reviews, and social media posts form a collective reference system that mediates the experiences of other travelers (Pan et al., 2007; Munar & Jacobsen, 2014).

Methodologically, UGC develops from a qualitative narrative source to quantitative data through *a text mining* and *sentiment analysis* approach (Pang & Lee, 2008). Text analytics allows the extraction of traveler's perception patterns at scale, including the analysis of hotel and destination experiences (Xiang et al., 2015). A big data-based approach is even used to predict tourism demand using online search data (Önder & Gunter, 2016).

A systematic review in the hospitality domain shows that the application of machine learning and text analytics has become the mainstream of modern tourism research (Mehraliyev et al., 2021). In addition, recent bibliometric studies confirm the exponential increase in the use of machine learning in tourism and hospitality research (Machine learning in tourism and hospitality, 2022).

Sentiment interpretation faces several significant challenges that affect the accuracy of detection, primarily related to linguistic complexity, contextual and cultural influences, as well as data and technical limitations. From a linguistic perspective, sentiment analysis models often struggle to capture implicit nuances such as sarcasm, irony, metaphors, and poorly structured sentences. In terms of context and culture, sentiment is highly shaped by situational background and socio-cultural environments, while lexicon-based approaches frequently overlook contextual meaning and variations in emotional expression across cultures. Technically, additional challenges arise from noisy user-generated data containing typographical errors, grammatical inconsistencies, and colloquial language, as well as potential bias in training datasets, domain dependency of models, and the absence of universal sentiment markers applicable across languages and disciplines (Aftab et al., 2023).

Sentiment analysis plays a significant role in destination decision-making processes. The analysis of tourist sentiment enables the measurement of perceived destination quality and provides strategic insights for sustainable development, including the identification of factors influencing prospective travelers' decisions based on online reviews and feedback. The results of comment analysis can function as a decision-support tool in destination management, assisting stakeholders in formulating evidence-based strategies. Furthermore, detailed sentiment analysis of online reviews can enhance the accuracy of tourism demand forecasting, which serves as an indicator of tourists' behavioral intentions toward specific destinations. This demonstrates the utility of sentiment as an input variable in decision-making and predictive models. Additionally, sentiment analysis supports destination managers in understanding how public opinion reflects and constructs perceptions of destination quality (Borrajo-Millán et al., 2021; Li et al., 2023; Tuhuteru et al., 2024).

Most sentiment analysis research in tourism focuses on text-based platforms such as TripAdvisor or Twitter (Xiang & Gretzel, 2010). However, the dynamics of social media show a significant shift towards video-based platforms. Travel videos allow for more complex spatial, emotional, and narrative representations than written text (Tussyadiah & Fesenmaier, 2009).

Platforms like YouTube form public discursive spaces that contain not only visual content, but also interactive commentary as a collective response to the destination's experience. Analysis of visual content and social media interactions shows that users not only consume information but form a community of shared perceptions (Hu et al., 2014). In the context of Indonesian tourism, studies on digital engagement in destination vlog reviews are starting to emerge, although they are still limited (Singgalen, 2025).

However, systematic exploration of trending YouTube vlog comments as a quantitative data source for destination sentiment analysis is still relatively rare compared to TripAdvisor-based research or other text-based platforms.

In this context, Tanah Lot represents a particularly relevant case for YouTube-based sentiment analysis. As one of Bali's most visually iconic coastal temple destinations, Tanah Lot is frequently featured in travel vlogs due to its dramatic sunset landscape, silhouette temple architecture, and ritual atmosphere. Its tourism appeal is strongly image-driven and highly mediated through video content, making YouTube an appropriate discursive arena to capture collective visitor perceptions. The dominance of visual storytelling in Tanah Lot's representation suggests that online commentary may reflect not only experiential evaluation but also mediated emotional responses shaped by digital imagery.

Previous studies indicate that class imbalance not only reduces model performance in general machine learning applications but also has a substantial impact on tourism data analysis, particularly in the sentiment classification of tourist reviews. In such contexts, minority classes such as negative reviews are often underrepresented, leading to biased predictive outcomes. To address this issue, oversampling techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) have been shown to improve the predictive accuracy of minority classes, thereby enhancing the robustness and reliability of sentiment classification models in tourism research. Moreover, User-Generated Content (UGC) as a primary data source in digital tourism studies has been extensively examined in relation to its influence on tourist decision-making and destination image formation. However, the literature review also highlights issues of limited inclusivity and content representation, which may be interpreted as a form of data imbalance within UGC-based tourism research. Such imbalance can affect the comprehensiveness and validity of findings, as certain voices, perspectives, or demographic groups may be overrepresented while others remain underrepresented in the dataset (Afrianto Singgalen et al., 2024; Sujatmiko et al., 2025).

In the realm of sentiment analysis, the Support Vector Machine (SVM) has long been used as a method of classification of high-dimensional texts due to its margin stability and computational efficiency (Pang & Lee, 2008). On the other hand, the development of deep learning, particularly Long Short-Term Memory (LSTM), offers the ability to capture more complex sequential dependencies and linguistic contexts (Li et al., 2018).

The application of machine learning in travel reviews has been carried out in various contexts, including travel products and hotels (Marrese-Taylor et al., 2013; Puh & Bagić Babac, 2023). Research in Indonesia has also begun to examine the comparison of SVM and LSTM in the sentiment analysis of tourist destinations (Ramdani & Cahyana, 2025), as well as the optimization of data balancing techniques such as SMOTE in the context of local tourism (Anggrawan et al., 2025). Systematic literature studies in the context of coastal tourism have also shown an increase in interest in machine learning-based sentiment analysis (Kurniawan et al., 2025).

In addition, research on the influence of the quality of social media information on visitor intent confirms that digital perception has direct implications for tourist behavior (Wang & Yan, 2022). The study of the influence of UGC on tourist visit intentions also strengthens the relevance of sentiment analysis in the context of destination management (Madaniah et al., 2024).

Based on the synthesis of the literature, there are four main research gaps:

1. The Dominance of Traditional Text Platforms

Most sentiment analysis studies still focus on TripAdvisor and Twitter (Xiang & Gretzel, 2010; Xiang et al., 2015), while systematic exploration of trending YouTube vlog comments as quantitative data is still limited.

2. Limitations of Aggregate Performance Evaluation

Although comparative studies of machine learning algorithms in tourism sentiment analysis are increasingly common, many of them rely predominantly on aggregate performance metrics such as accuracy or overall F1-score (Mehraliyev et al., 2021; Machine learning in tourism and hospitality, 2022). While these metrics provide a general indication of model performance, they often obscure class-level behavior, particularly in datasets characterized by imbalance. As a result, models may appear highly accurate while systematically failing to capture minority sentiment categories that are critical for interpretation.

3. Methodological Urgency

The growing reliance on aggregate evaluation metrics without examining class-wise performance introduces a critical methodological limitation in comparative sentiment analysis. In imbalanced data contexts, marginal differences in accuracy or weighted metrics can mask substantial disparities in how models represent minority classes. Consequently, model evaluation that is not grounded in detailed error analysis and class-level performance risks overestimating the practical effectiveness of certain algorithms.

From an academic perspective, this limitation reduces the interpretability and comparability of findings, as conclusions are drawn from metrics that do not fully reflect underlying classification behavior. More importantly, from a managerial standpoint, such oversights can lead to distorted representations of public perception. In the context of destination governance, sentiment analysis functions as a decision-support mechanism for identifying visitor perceptions. If models fail to adequately capture minority sentiments, such as negative feedback, destination managers may overlook critical signals, resulting in incomplete or biased policy responses. Therefore, there is a methodological need to shift from purely aggregate evaluation toward a more granular analysis of model behavior, particularly under conditions of limited and imbalanced tourism data.

4. Context of Bali's Specific Destinations

Bali, as a mature international tourism destination with highly developed infrastructure and strong global branding, faces complex visitor perceptions shaped by rising overtourism, environmental pressure, and cultural commodification. Simultaneously positioned as both a spiritual icon and a commercial lifestyle hub, its destination image is further amplified by social media and user-generated content (UGC), which intensify visitor concentration in key hotspots. In this context, Bali provides a relevant setting to compare Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) models in analyzing multimodal video UGC to better understand how tourists construct meanings of spirituality and commercialization amid overtourism pressures (Hm & Hulu, 2025; Ningsih & Fachrurreza, 2020; Sudiarta et al., 2025).

Within Bali's tourism ecosystem, Tanah Lot occupies a strategic and symbolic position. Beyond functioning as a high-volume tourist attraction, it is simultaneously an active Hindu temple and a sacred coastal landscape. This dual identity sacred space and mass tourism commodity creates a complex perception structure that may generate polarized or ambivalent sentiment expressions in digital discourse. The coexistence of religious ritual, sunset crowd concentration, commercial activity, and global branding makes Tanah Lot an analytically rich site for examining how visitors negotiate meanings of spirituality, authenticity, congestion, and commodification in online commentary. Despite its symbolic importance, quantitative sentiment analysis focusing specifically on Tanah Lot using statistically validated model comparison remains limited.

As tourism management increasingly relies on automated analytics, the reliability of algorithmic interpretation becomes not merely a technical concern but a governance issue. Therefore, statistical validation of model differences is essential to ensure that managerial interpretations of digital sentiment are grounded in robust analytical foundations.

Thus, this study seeks to conduct a comparative evaluation of SVM and LSTM on YouTube vlog comment data for Tanah Lot destinations in a span of one year. This approach not only makes a methodological contribution to the analysis of tourism sentiment but also provides a quantitative basis for the managerial interpretation of destinations. Beyond statistical comparison, this study also seeks to interpret the structural characteristics of sentiment expression within YouTube's discursive environment.

While deep learning architectures such as LSTM are typically associated with large-scale datasets, tourism sentiment research frequently operates under limited data conditions, particularly when focusing on specific destinations, short temporal windows, or platform-specific discourse. This study deliberately evaluates deep learning performance under low-resource tourism data conditions to examine whether sequential architectures maintain analytical robustness when applied to small, imbalanced, and context-specific user-generated content. Such a scenario reflects realistic constraints faced by destination managers who rely on niche or localized digital datasets.

Based on the background and research gaps that have been identified, this research is formulated in the following questions:

1. The characteristics of the distribution of sentiment comments on trending YouTube vlogs regarding Tanah Lot tourist destinations in a span of one year have not been identified quantitatively and systematically.
2. The performance of sentiment classification using the Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) models on YouTube comment data to Tanah Lot needs to be evaluated and compared based on measurable evaluation metrics.
3. The significance of the difference in performance between the SVM and LSTM models in sentiment classification has not been tested using a formal inferential statistical approach.
4. The misclassification structure of the two models and their implications for sentiment interpretation in the context of tourist destination management have not been comprehensively analyzed.

In line with the formulation of the problem, this research aims to:

1. Analyzing the distribution of sentiment on YouTube vlogs of Tanah Lot destinations as a representation of digital public perception of tourist destinations.
2. Evaluate and compare sentiment classification performance using machine learning (SVM) and deep learning (LSTM) approaches.

E-Jurnal Matematika Vol. 15(2), Mei 2026, pp. 48-58 ISSN: 2303-1751
DOI: https://doi.org/10.24843/MTK.2026.v15.i02.p502

3. Identify patterns of misclassification to provide a quantitative basis in the managerial interpretation of tourist destinations.

2. RESEARCH METHODS

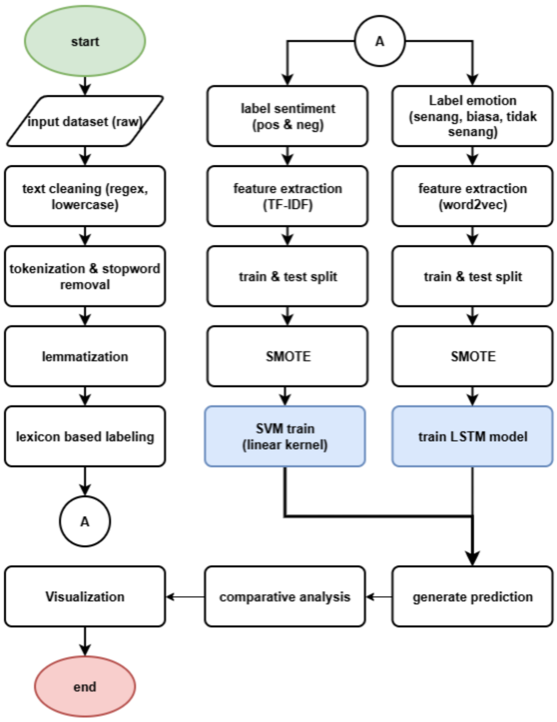

Figure 2. Research method flowchart

This study uses a comparative experimental design to compare the performance of Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) in classifying the sentiment of YouTube comments related to Tanah Lot tourist destinations. The process starts from collecting comment data (raw text) which then goes through pre-processing stages in the form of text cleaning (regex and lowercase), tokenization, stopword removal, and lemmatization. Labeling is done automatically using a lexicon-based labeling approach. The dataset was processed using a unified three-class sentiment classification scheme consisting of Positive, Neutral, and Negative categories. Both SVM and LSTM models were trained and evaluated using the same labeled dataset to ensure comparability of predictions and class imbalance is handled in the training data using SMOTE.

In the SVM pathway, text is represented using TF-IDF before training with a linear kernel. In the LSTM pathway, text is represented in the form of word embedding, one-hot encoding is carried out on the label, and input dimension adjustment before model training with a single-layer LSTM architecture and softmax-activated dense layer. The evaluation of the two models was carried out using accuracy, precision, recall, F1-score, and confusion matrix. To test the significance of the difference in performance between the two models. Datasets are defined as:

$$D = \{(x_i, y_i)\}_{i=1}^{n}$$

with $x_i$is the first and most important comment. $y_i$auto-labelling sentiment labels. Data was obtained from YouTube comments related to Tanah Lot in the span of one year.

The data is divided into:

a. 80% of training data.

b. 20% of test data.

Significant class imbalances were identified, with neutral class dominance. The dataset manually cleaned and labeled comments. The study intentionally maintains this scale to simulate a low-resource tourism sentiment scenario, where destination-specific discourse is often limited in volume but high in contextual richness. The inclusion of LSTM in this experimental design serves as a benchmarking exercise to evaluate whether deep learning architectures offer performance advantages over classical linear models under constrained data conditions.

While deep learning architectures such as LSTM are typically associated with large-scale datasets, tourism sentiment research frequently operates under limited data conditions, particularly when focusing on specific destinations, short temporal windows, or platform-specific discourse. This study deliberately evaluates deep learning performance under low-resource tourism data conditions to examine whether sequential architectures maintain analytical robustness when applied to small, imbalanced, and context-specific user-generated content. Such a scenario reflects realistic constraints faced by destination managers who rely on niche or localized digital datasets.

2.1. Support Vector Machine

The SVM model uses a linear kernel with the decision function:

$$f(x) = w^T x + b$$

with optimization:

$$\min_{w,b} \frac{1}{2} || w ||^2 + C \sum_{i=1}^{n} \xi_i$$

The feature is represented using TF-IDF with the following parameters:

a. max_features = 5000,
b. min_df = 5,
c. Unigram.

2.2. Long Short-Term Memory

The LSTM architecture course consists of:

a. LSTM layer (100 unit),
b. Dense layer (50 unit, ReLU),
c. Output layer (Softmax).

Mathematically, the mechanism of LSTM follows the equation:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$
$$C_t = f_t C_{t-1} + i_t \tilde{C}_t$$
$$h_t = o_t \tanh(C_t)$$

As much as 10% of the training data is used as an internal validation set.

2.3. Model Evaluation

Evaluation metrics include:

Accuracy:

$$\frac{TP + TN}{TP + TN + FP + FN}$$

Precision:

$$\frac{TP}{TP + FP}$$

Recall:

$$\frac{TP}{TP + FN}$$

F1 Score:

$$2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

instances with identical categorical outcomes. Therefore, sentiment labels were standardized prior to model training to avoid label-space discrepancies that could invalidate inferential comparison.

3. RESULTS AND DISCUSSION

3.1. Result

LSTM Model Performance Metrics (classification_report from previous step):

	precision	recall	f1-score	support
	0.95	1.00	0.97	19
	0.00	0.00	0.00	1
accuracy			0.95	20
macro avg	0.47	0.50	0.49	20
weighted avg	0.90	0.95	0.93	20

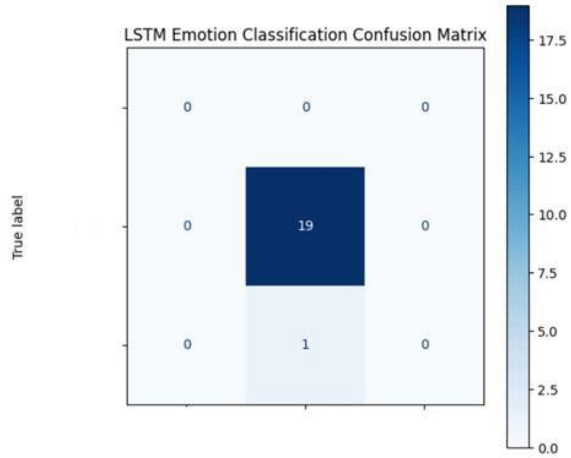

Figure 3.1 LSTM Model Performance Metric
Source: Researcher's Analysis (2026)

SVM Model Performance Metrics (from previous step):
Accuracy: 0.9500
Precision (weighted): 0.9025
Recall (weighted): 0.9500
F1-Score (weighted): 0.9256

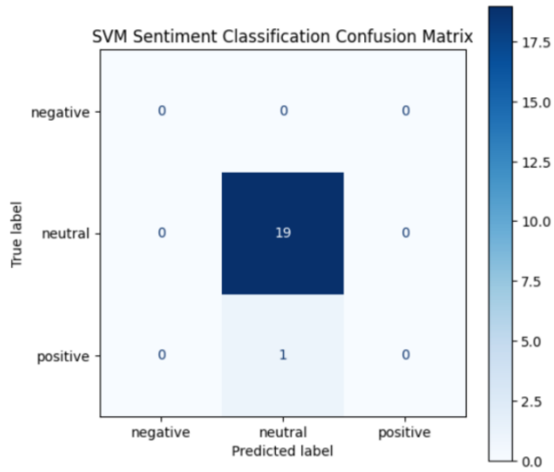

Figure 3.2 SVM model performance metric
Source: Researcher's Analysis (2026)

The results of the evaluation showed that both models, SVM and LSTM, produced an equally high accuracy value of 0.95 on the test data. The SVM model obtained a weighted precision of 0.9025, a recall of 0.9500, and an F1-score of 0.9256. Meanwhile, the LSTM model also showed an accuracy of 0.95 with high precision, recall, and F1-score in the majority ("mediocre") class. Descriptively, both models appear to have excellent performance. However, the confusion matrix analysis showed that the high accuracy was dominated by the success of classification in the majority classes, namely neutral (SVM) and ordinary (LSTM), which indeed dominated the distribution of test data.

In the SVM model, all 19 neutral data in the test data were correctly classified. However, the model does not manage to classify positive and negative classes consistently. This shows the model's tendency to predict the majority class when the distribution of data is unbalanced. In the LSTM model, more extreme patterns are identified. The model almost exclusively predicts ordinary classes, with a recall of 0.00 for both positive and unpositive classes. In other words, despite its high accuracy, the LSTM model fails to capture the minority class and exhibits a strong bias against the majority class. This phenomenon is consistent with the characteristics of datasets that have significant class imbalances.

Comparative analysis of the first 10 comments shows a systematic pattern of divergence. Some comments that were predicted to be positive by the SVM were still categorized as mediocre by the LSTM. Even on comments that contain explicit expressions such as "Good" or "Like", LSTM does not classify as positive. There was not a single case in which the two models simultaneously identified the positive/neutral category. This suggests that SVMs still have limited sensitivity to keyword-based positive signals, while LSTMs tend to generalize predictions to dominant classes due to unbalanced data distribution learning.

The main findings of this study can be summarized as follows: (1) both models show high accuracy which is influenced by the dominance of the majority class; (2) LSTM has a strong bias against ordinary classes and fails to identify minority classes; (3) SVM shows limited ability to detect positive classes even though it remains biased towards neutral; and (4) class imbalance is the dominant factor that affects the classification results.

Substantively, these results suggest that in the context of YouTube-based tourism sentiment data with limited numbers and unbalanced distribution, the selection of the model should take into account stability towards the majority class and sensitivity to the minority class. Accuracy-based evaluation alone is not enough to assess model performance, so statistical testing and error distribution analysis are crucial in the interpretation of results.

3.2. Discussion

The results of this study provide empirical evidence regarding the comparative behavior of traditional machine learning models (SVM) and deep learning architectures (LSTM) within the context of limited and imbalanced tourism data. In this study, sentiment labels are defined consistently across both learning methods. Each comment is assigned to one of three sentiment categories: Positive, Neutral, or Negative. For the LSTM model, the term "emotion label" refers to the same underlying sentiment category, with the designation used internally to process one-hot encoded labels for deep learning. Although SMOTE is applied in the LSTM pathway as part of the model training augmentation, the fundamental sentiment labels do not change; they remain strictly Positive, Neutral, or Negative.

Therefore, the sentiment variable is identical and consistent for both SVM and LSTM, serving as a common reference for performance evaluation. The distinction between methods arises from how each model learns and predicts these labels: SVM relies on feature-based linear separation with TF-IDF representations, whereas LSTM learns sequential patterns from word embeddings. This ensures that the comparison between methods is valid and directly comparable at the instance level. Although both models yielded identical accuracy values (0.95), this similarity does not reflect equivalent classification performance. The findings demonstrate that reliance on a single aggregate metric, such as accuracy, can obscure substantial differences in how models handle class distribution, particularly in datasets dominated by a majority class.

A more detailed class-wise evaluation reveals that the LSTM model exhibited a strong bias toward the majority class (neutral), resulting in a recall of 0.00 for minority classes. This indicates that the model failed to identify both positive and negative sentiments, despite achieving high overall accuracy. Such behavior suggests that the model minimizes loss by consistently predicting the dominant class, rather than learning meaningful sentiment distinctions.

This outcome highlights a critical limitation of deep learning architectures such as LSTM in low-resource scenarios. LSTM models are inherently data-intensive, requiring large volumes of training data to effectively capture sequential linguistic patterns. When trained on a small dataset of only 300 samples, the model lacks sufficient exposure to diverse sentiment structures, leading to poor generalization and class collapse toward the majority category.

In contrast, the SVM model demonstrated more stable performance across sentiment classes. As a feature-based approach, SVM does not rely on sequential pattern learning but instead operates on structured representations of text, making it more robust under conditions of limited and imbalanced data. This difference in learning mechanisms explains why SVM was able to retain sensitivity to minority classes while LSTM failed to do so.

These findings reinforce the importance of context-aware model selection in tourism sentiment analysis. The assumption that more complex models inherently yield better performance is not supported under constrained data conditions. Instead, model effectiveness should be evaluated based on its ability to capture meaningful class distinctions, particularly when minority sentiments such as negative feedback, carry significant implications for destination management.

From a practical perspective, the inability of a model to detect minority sentiments poses a substantial risk for decision-making processes. In the context of tourism governance, overlooking negative sentiment may lead to an incomplete understanding of visitors experience and delay necessary interventions. Therefore, evaluation strategies should prioritize class-wise performance metrics to ensure that analytical outputs remain aligned with managerial needs.

Overall, this study demonstrates that comparative evaluation must move beyond aggregate metrics toward a more granular analysis of model behavior, particularly in real-world tourism datasets characterized by limited size and class imbalance.

3.2.1 Model Stability Comparison in Limited Data Scenarios

The phenomenon in which SVM exhibits relatively superior stability compared to LSTM on a small dataset confirms the well-established dependency of sequential architectures on large-scale training data, where insufficient sample size leads to representational collapse toward the majority class, the LSTM almost exclusively predicted the majority class ('neutral') and failed entirely to identify minority classes ('positive/negative'), resulting in a recall value of 0.00. Conversely, the SVM, supported by TF-IDF feature representation remained capable of capturing positive keyword signals, albeit with limited efficacy. This aligns with the argument that linear models are often more robust in low-resource tourism sentiment scenarios, as they do not require the complex parameter tuning inherent in sequential architecture.

3.2.2 Implications for Tanah Lot Destination Governance

From a destination management perspective, the inability of the models (particularly LSTM) to detect specific negative or positive sentiments has serious implications for data-driven decision-making. Tanah Lot, as a tourism icon with a dual identity as both a sacred site and a mass tourism commodity, generates a complex perceptual structure on YouTube. Should destination managers rely solely on models biased toward neutral sentiment, tourist grievances regarding traffic congestion, cultural commodification, or environmental degradation will remain undetected by automated systems.

3.2.3 The Challenge of Class Imbalance

From a destination management perspective, the inability of the models (particularly LSTM) to detect specific negative or positive sentiments has serious implications for data-driven decision-making. Tanah Lot, as a tourism icon with a dual identity as both a sacred site and a mass tourism commodity, generates a complex perceptual structure on YouTube. Should destination managers rely solely on models biased toward neutral sentiment, tourist grievances regarding traffic congestion, cultural

commodification, or environmental degradation will remain undetected by automated systems.

The results of this study directly address the research question regarding the comparative effectiveness of SVM and LSTM in sentiment classification on a small-scale dataset. The findings indicate that while both models achieved high accuracy, their functional effectiveness remains significantly disparate. SVM proved more capable of addressing classification challenges within imbalanced datasets by successfully capturing positive signals, whereas LSTM failed to meet the classification objectives by merely predicting the majority class. This confirms that for research problems focused on low-resource conditions, such as those involving tourist destinations like Tanah Lot, linear models utilizing TF-IDF features provide a more reliable solution than complex deep learning architectures. This pattern confirms that high accuracy in imbalanced datasets may primarily reflect correct predictions of the dominant class rather than true classification capability.

Consequently, the superior performance of Support Vector Machines (SVM) over the stability of Long Short-Term Memory (LSTM) networks in this study ensures that digital data from YouTube can be effectively converted into precise managerial insights, thereby supporting sustainability strategies for this iconic Balinese destination.

4. CONCLUSIONS AND SUGGESTIONS

The results of this study demonstrate that while both SVM and LSTM achieved identical accuracy (0.95) on the limited and imbalanced Tanah Lot YouTube dataset, their class-wise performance differs substantially: LSTM failed to detect minority classes (Positive and Negative), producing a recall of 0.00, whereas SVM maintained partial sensitivity due to its feature-based linear approach. This highlights a critical limitation of deep learning architectures such as LSTM in low-resource scenarios, where insufficient data prevents effective sequential learning and leads to majority-class bias, while classical models like SVM remain robust under such constraints. The sentiment labels, Positive, Neutral, and Negative are consistent across both methods, with the “emotion label” in LSTM serving as an internal representation without altering the underlying data, ensuring valid comparisons. These findings underscore the importance of evaluating model behavior beyond aggregate metrics and emphasize context-aware model selection, particularly in tourism sentiment analysis where detecting minority sentiments is vital for destination governance.

For future research, larger and more balanced datasets, advanced imbalance mitigation techniques, and multi-platform or multimodal data sources are recommended to enhance minority-class detection and generalize model performance, while evaluation should continue to prioritize class-wise metrics and detailed error analysis to ensure interpretability and managerial relevance.

REFERENCES

Anggrawan, A., Hairani, H., & Satria, C. (2025). Optimizing sentiment analysis for Lombok tourism using SMOTE and chi-square with machine learning. *RESTI Journal (Systems Engineering and Information Technology), 9*(4), 706–713. https://doi.org/10.29207/resti.v9i4.6623

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. *Tourism Management, 29*(4), 609–623. https://doi.org/10.1016/j.tourman.2008.01.005

Dang, T. D., & Nguyen, M. T. (2023). Systematic review and research agenda for the tourism and hospitality sector: Co-creation of customer value in the digital age. *Future Business Journal, 9*, 94.

Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: Foundations and developments. *Electronic Markets, 25*(3), 179–188. https://doi.org/10.1007/s12525-015-0196-8

Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we Instagram: A first analysis of Instagram photo content and user types. *Proceedings of the International AAAI Conference on Web and Social Media, 8*(1), 595–598. https://doi.org/10.1609/icwsm.v8i1.14578

Kim, W. G., Li, J. J., & Brymer, R. A. (2016). The impact of social media reviews on restaurant performance: The moderating role of excellence certificate. *International Journal of Hospitality Management, 55*, 41–51. https://doi.org/10.1016/j.ijhm.2016.03.001

Kindzule-Millere, I., & Zeverte-Rivza, S. (2022). Digital transformation in tourism: Opportunities and challenges. *Economic Science for Rural Development, 56*, 476–486.

Kurniawan, S., Pramayoga, A. S., & Ashari, Y. F. (2025). Sentiment analysis in coastal tourism: A systematic literature review. *Engineering, Technology & Applied Science Research, 15*(6), 30226–30233. https://doi.org/10.48084/etasr.14644

Law, R., Buhalis, D., & Cobanoglu, C. (2014). Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management, 26*(5), 727–750. https://doi.org/10.1108/IJCHM-08-2013-0367

Li, Q., Li, S., Hu, J., Zhang, S., & Hu, J. (2018). Tourism review sentiment classification using a bidirectional recurrent neural network with an attention mechanism and topic-enriched word vectors. *Sustainability, 10*(9), 3313. https://doi.org/10.3390/su10093313

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management, 29*(3), 458–468. https://doi.org/10.1016/j.tourman.2007.05.011

Machine learning in tourism and hospitality: A bibliometric analysis and research agenda. (2022). *International Journal of Hospitality Management, 107*, 103317. https://doi.org/10.1016/j.ijhm.2022.103317

Madaniah, S., Situmorang, S. H., & Sembiring, B. K. F. (2024). Influence of user-generated content on tourist visit intention: A literature review. *Jurnal Mantik, 8*(3), 156–172. https://doi.org/10.35335/mantik.v8i3.5751

Marrese-Taylor, E., Velásquez, J. D., & Bravo-Marquez, F. (2013). A novel deterministic approach for aspect-based opinion mining in tourism products reviews. *Expert Systems with Applications, 41*(17), 7764–7775. https://doi.org/10.1016/j.eswa.2014.05.022

Mehraliyev, F., Chan, I. C. C., Kirilenko, A. P., & Liu, Y. (2021). Sentiment analysis in hospitality and tourism: A thematic and methodological review. *International Journal of Contemporary Hospitality Management, 33*(1), 46–77. https://doi.org/10.1108/IJCHM-02-2021-0132

Munar, A. M., & Jacobsen, J. K. S. (2014). Motivations for sharing tourism experiences through social media. *Tourism Management, 43*, 46–54. https://doi.org/10.1016/j.tourman.2014.01.012

Önder, I., & Gunter, U. (2016). Forecasting tourism demand with Google Trends: Accuracy comparison of countries versus cities. *International Journal of Tourism Research, 18*(4), 297–306. https://doi.org/10.1002/jtr.2046

Pan, B., MacLaurin, T., & Crotts, J. C. (2007). Travel blogs and the implications for destination marketing. *Journal of Travel Research, 46*(1), 35–45. https://doi.org/10.1177/0047287507302378

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135. https://doi.org/10.1561/1500000011

Puh, K., & Bagić Babac, M. (2023). Predicting sentiment and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights, 6*(3), 1188–1204. https://doi.org/10.1108/JHTI-02-2022-0078

Ramdani, R., & Cahyana, R. (2025). Cultural tourism review sentiment analysis: Comparison of SVM and LSTM methods. *Journal of Algorithms, 22*(2), 1188–1199. https://doi.org/10.33364/algoritma.v22i2.2585

Sigala, M., Christou, E., & Gretzel, U. (2012). Social media in travel, tourism and hospitality: Theory, practice and cases. *International Journal of Contemporary Hospitality Management, 24*(3), 436–438. https://doi.org/10.1108/09596111211217991

Singgalen, Y. A. (2025). Understanding digital engagement through sentiment analysis of tourism destination travel vlog reviews. *KLIK: Scientific Studies of Informatics and*

Computers, 4(6). https://doi.org/10.30865/klik.v4i6.1947

Tussyadiah, I. P., & Fesenmaier, D. R. (2009). Mediating tourist experiences: Access to places via shared videos. *Annals of Tourism Research, 36*(1), 24–40. https://doi.org/10.1016/j.annals.2008.10.001

Wang, D., Li, X., & Li, Y. (2013). China's "smart tourism destination" initiative: A taste of the service-dominant logic. *Journal of Destination Marketing & Management, 2*(2), 59–61. https://doi.org/10.1016/j.jdmm.2013.05.004

Wang, H., & Yan, J. (2022). Effects of social media tourism information quality on destination travel intention: The mediating role of self-congruity and trust. *Frontiers in Psychology, 13*, 1049149. https://doi.org/10.3389/fpsyg.2022.1049149

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management, 31*(2), 179–188. https://doi.org/10.1016/j.tourman.2009.02.016

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience? *International Journal of Hospitality Management, 44*, 120–130. https://doi.org/10.1016/j.ijhm.2014.10.013

Zhang, Z., Ye, Q., Law, R., & Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management, 29*(4), 694–700. https://doi.org/10.1016/j.ijhm.2010.02.002