
Available online at https://journals.sjp.ac.lk

**JOURNAL OF REAL ESTATE STUDIES**

ISSN 3051-4878 [PRINT] | E-ISSN 3051-4886 [ONLINE]

# A Machine Learning-Based Sentiment Analysis of Online International Tourist Reviews: A Study of Heritage Sites in Anuradhapura, Sri Lanka

**Hashini T. Wickremasinghe** [a*]

[a] *Deakin University, Australia*

## ABSTRACT

In the contemporary digital landscape, online reviews and other user-generated content significantly influence global travel behaviours by revealing first-hand visitor perspectives. However, these insights remain largely underexplored in tourism research and destination management. This study examines the application of text mining and machine learning-based sentiment classification to analyse tourist reviews of heritage sites in Anuradhapura, Sri Lanka, aiming to produce data-driven insights to support tourism planning. Reviews collected from Google and TripAdvisor between 2018 and 2024 were pre-processed and categorized into positive, neutral, or negative sentiments using a decision tree classifier. The model produced an overall accuracy of 80.85% with substantial inter-rater agreement (kappa = 0.634), effectively identifying prevailing sentiment patterns. Findings show that positive sentiments were driven by aesthetic value, architectural heritage, and spiritual enrichments, while negative sentiments emphasized operational deficiencies, unmet visitor expectations, and disappointing on-site conditions. The study highlights the value of automated sentiment analysis as a decision-support tool for heritage site management, offering actionable insights for targeted improvements and enhancing visitor satisfaction. The proposed approach is scalable and transferable to other heritage destinations worldwide.



## ARTICLE INFO

***Article History:***

Received 20th December 2025
Revised 17th January 2026
Accepted 25th January 2026

***Keywords:***

Text Mining, Sentiment Analysis, Natural Language Processing, Machine Learning, Tourism Analytics, Sri Lanka

## 1. INTRODUCTION

In the modern digital landscape, online reviews and user-generated content have become powerful forces shaping travel choices across the globe, offering abundant but often untapped insights into visitors' experiences. Platforms like TripAdvisor, Google Reviews, and social media host a rapidly growing volume of traveller feedback, offering valuable experiential data about tourism destinations (Mutalib et al., 2021). However, much of this data is unstructured and conveyed in natural

* Corresponding author. Tel: +61452399547 ; Email: s224932797@deakin.edu.au | hashinitw@gmail.com; *https://orcid.org/0000-0003-0591-3477*
Institution: Deakin University, Australia
Doi: 10.31357/jres.v23i01.9034

language, making it challenging to analyze using traditional methods. Conventional approaches such as surveys or field interviews, are constrained by small sample sizes, resource-intensive, limited in scope, single-dimensional data, and subjective biases, making it difficult to comprehensively capture visitors' fine-grained perceptions in diverse scenarios and often fail to capture the dynamic, real-time sentiments expressed by tourists online (Yuan et al., 2025). Consequently, tourism managers and policymakers lack comprehensive, timely insights necessary for effective decision-making.

Tourism is an important part of Sri Lanka's economy, especially in the heritage tourism sector. Despite the cultural and economic significance of sites like Anuradhapura, the country's ancient UNESCO World Heritage city, modern data analytics have been used only minimally to understand visitor perceptions systematically. Harnessing advanced computational methods, particularly sentiment classification within the field of Natural Language Processing (NLP), offers the potential to transform vast amounts of online review data into actionable insights. By automatically identifying positive, neutral, or negative sentiments, such techniques enable a deeper understanding of visitor experiences and satisfaction drivers at scale.

This study aims to develop and evaluate text mining-based sentiment classification models on tourist reviews of Anuradhapura's heritage sites for data-driven tourism management. The goal is to uncover key terms and areas for improvement that may otherwise go unnoticed in traditional analysis. The findings will provide tourism authorities, local policymakers, destination managers, and digital marketers with data-driven guidance to enhance service quality, preserve cultural heritage, and optimize visitor experiences.

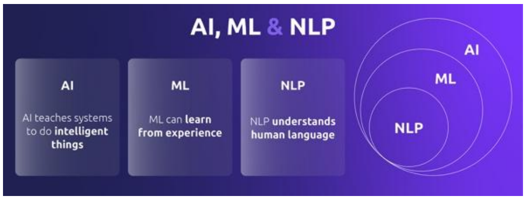
## 2. LITERATURE REVIEW

### 2.1. Artificial Intelligence, Machine Learning and Natural Language Processing

Artificial Intelligence (AI) encompasses computational techniques designed to emulate human cognitive functions such as reasoning, learning, perception, and decision-making (Kok, Boers & Kosters, 2009). Within AI, Machine Learning (ML) constitutes a major methodological paradigm focused on developing algorithms capable of detecting patterns from data (Zhou, Song, Ji, & Wei, 2022) and improving performance through experience rather than explicit programming. ML algorithms learn statistical representations from structured or unstructured data, enabling predictive modelling, classification, and pattern recognition tasks at scale.

Natural Language Processing (NLP) represents a specialized subfield of AI concerned with enabling machines to interpret, process, and generate human language in both written and spoken form (Cambria & White, 2014). NLP integrates computational linguistics with ML and deep learning to address tasks such as syntactic parsing, semantic interpretation, named entity recognition, topic modelling, text classification, and language generation. Contemporary NLP techniques are often underpinned by Large Language Models (LLMs) and deep learning architectures, which leverage large-scale corpora and neural network models to capture contextual, semantic, and pragmatic nuances in language.

**Figure 1: AI, ML and NLP**



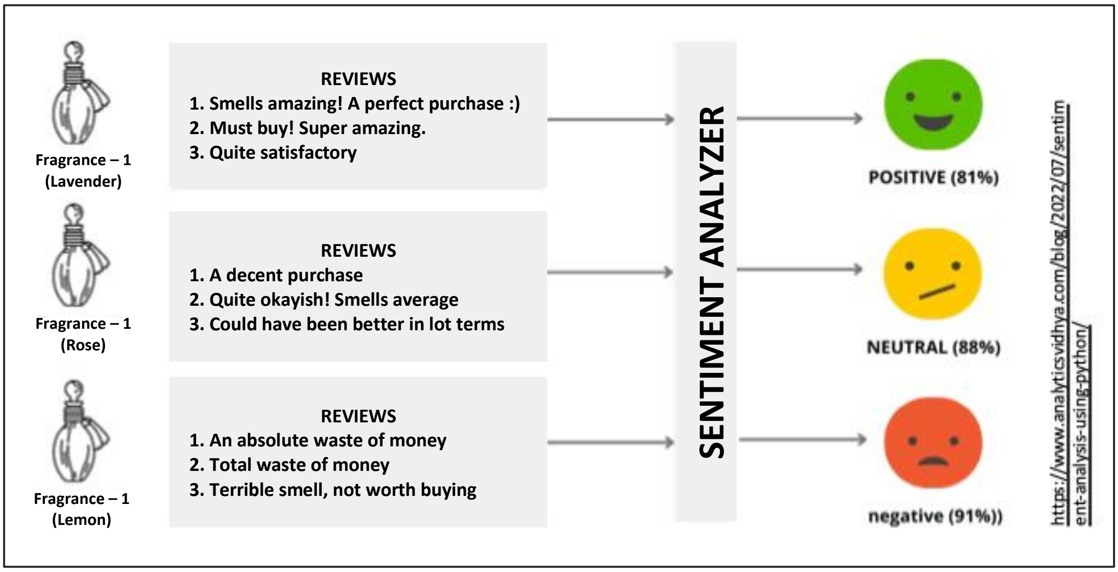

*Source: Diabolocom (2023)*

NLP has become increasingly central to text analytics and text mining, offering scalable mechanisms for extracting actionable knowledge from high-volume unstructured data sources such as reviews, social media posts, and digital communications. Figure 1 illustrates the hierarchical relationship between AI, ML, and NLP, positioning NLP as a domain that applies ML techniques to linguistic data. Through these advances, NLP contributes to numerous downstream applications, including information retrieval, conversational agents, translation systems, and sentiment analysis.

### 2.2. Sentiment Analysis (SA)

Sentiment Analysis (SA), also referred to as opinion mining, is a core NLP task aimed at identifying and categorizing affective states, attitudes, or evaluative judgments embedded within text (Yusof, Mohamed, & Abdul-Rahman, 2019).

**Figure 2: Sentiment Analysis**




*Source: Analytics Vidhya (2025)*

As shown in Figure 2, it typically involves the classification of textual units (e.g., sentences, reviews, or documents) into sentiment categories such as positive, negative, or neutral. Beyond polarity detection, advanced SA frameworks capture emotional intensity, stance, and subjectivity, enabling richer interpretation of public perceptions and behavioural responses.

In tourism research, SA has emerged as a significant methodological tool for analyzing visitor opinions and enhancing evidence-based destination management. The proliferation of user-generated content (UGC) on platforms such as TripAdvisor, Twitter (X), and Instagram has provided unprecedented access to authentic visitor narratives. SA enables tourism scholars and practitioners to examine how destinations are perceived, what attributes shape visitor satisfaction, and how these perceptions influence behavioural intentions and consumption patterns (Xiang et al., 2015; Surugiu et al., 2023). Such insights contribute to strategic marketing, experience optimisation, heritage conservation decisions, and resource allocation.

Recent advances in SA are driven by developments in ML and deep learning. While earlier approaches relied on lexicon-based models or hand-crafted linguistic features, contemporary studies increasingly employ supervised machine learning approaches and transformer-based architectures. Models such as BERT, RoBERTa, and XLNet enhance contextual representation by leveraging bidirectional encoding and attention mechanisms, substantially outperforming traditional classifiers in sentiment classification tasks (Viñan-Ludeña & de Campos, 2022). Studies in heritage tourism further demonstrate the applicability of SA for cultural asset evaluation: Gulati (2022) explored Twitter sentiments related to Indian heritage sites, while Yuan et al. (2025) analyzed public perceptions of architectural heritage to inform sustainable conservation strategies. Similarly, research by Singgalen (2023) and Mutalib et al. (2021) applied decision trees, support vector machines, and regression-based models to predict visitor behaviour and classify affective responses.

Tourist perception of cultural and environmental values has been widely examined in recent studies, with particular emphasis on emotional engagement. Sima et al. (2024) found that emotional attachment and awareness of conservation efforts significantly influence tourists' perceptions and satisfaction at coastal destinations. Similarly, Lin, Zhao, and Wang (2024), using sentiment analysis at Saguaro National Park, highlighted emotional value as a key determinant of tourist satisfaction. These findings indicate that tourists' perceptions are shaped not only by functional attributes but also by emotional connections with destinations. Research on cultural tourism further emphasizes this emotional dimension. Medina-Viruel et al. (2019) demonstrated that cultural motivations and emotional experiences are central to tourist satisfaction in World Heritage cities, underscoring the interplay between cognitive and affective factors in culturally significant destinations.

The growing use of social media has enabled new approaches to understanding tourist sentiment. Gandy et al. (2024) evaluated automated sentiment analysis tools, such as VADER, highlighting their effectiveness and limitations in tourism-related social media data. Additionally, Bhardwaj, Bharany, and Kim (2024) proposed a sentiment-based machine learning framework to distinguish genuine from misleading online content, offering important implications for tourism research where online sentiment strongly influences destination image and travel decisions.

Despite global progress, regional applications remain uneven. In South Asia, and particularly in Sri Lanka, there is limited empirical research applying advanced text mining and sentiment classification methods to heritage tourism contexts. Iconic cultural heritage sites such as Anuradhapura attract substantial foreign visitation, yet systematic analyses of visitor narratives remain scarce. Linguistic and cultural nuances embedded in visitor reviews also present modelling challenges that have yet to be examined in depth. Addressing this gap, the present study develops and evaluates sentiment classification models using foreign visitor reviews related to Anuradhapura's heritage sites, providing data-driven insights to support sustainable heritage management and strategic tourism planning.

## 3. RESEARCH METHODS

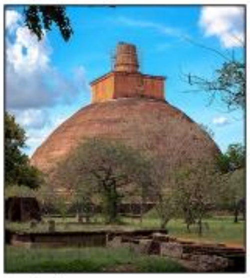
### 3.1. Study Area

This study analyzed tourist reviews of major heritage sites in Anuradhapura to develop sentiment classification models for data-driven tourism management. As shown in figures 3-10, the selected sites are Ruwanweli Maha Seya, Sri Maha Bodhiya, Jetavanaramaya, Isurumuniya Temple, Kuttam Pokuna (Twin Ponds), Mihintale, Abhayagiri Stupa and Ritigala Monastery. These were ranked as top visited places in Anuradhapura by both platforms of Google Reviews and TripAdvisor in years 2024 and 2025

**Figures 3-10: Selected Heritage Sites**

Sri Maha Bodhiya

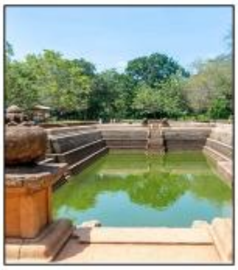
Ruwanweli Maha Seya (Stupa)



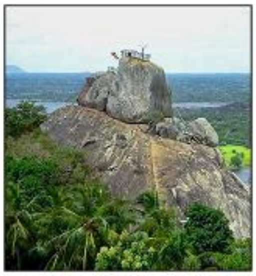
Jetavanaramaya (Stupa)

Abhayagiri Stupa

Isurumuniya Temple



Kuttam Pokuna (Twin Ponds)



Mihintale

Ritigala Monastery

*Source: Author (2025), Tripadvisor (2025), Story of Sri Lanka (2024), Punchihewa (2021), Travel Setu (2025)*

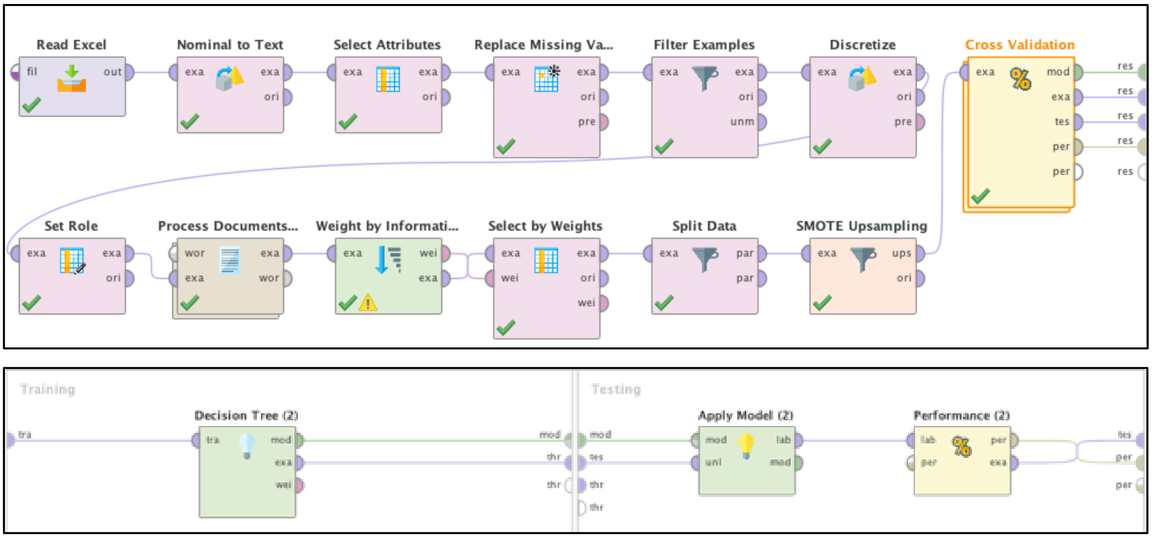
### 3.2. Methods

Reviews were collected from public platforms, specifically Google Reviews and TripAdvisor during January 2018 to December 2024 period, using the Instant Data Scraper tool. The reviews from foreigner visitors only were retained, and local visitors or irrelevant content was removed to ensure dataset quality and consistency.

The collected dataset underwent a comprehensive text mining process using the Altair AI Studio tool as shown in figure 11, to convert unstructured textual data into structured features suitable for modelling. Preprocessing included transforming text to lowercase, tokenization, stemming, stop word removal, filtering tokens by length, and noise elimination such as punctuation, numbers, and special characters. After preprocessing, a weigh-select dimensionality reduction method was applied to retain the most informative features.

To create sentiment labels for supervised learning, the star ratings provided by reviewers were discretized based on upper limits: Positive ($\leq$5.0), Neutral ($\leq$3.0), and Negative ($\leq$2.0), allowing textual content to be mapped consistently to sentiment categories. The dataset was then split into training and testing sets using a 70:30 ratio. To address class imbalance, SMOTE upsampling was applied to the training set, ensuring that minority classes were adequately represented.

**Figure 11: Workflow for Sentiment Classification of Tourism Reviews**




*Source: Author (2025)*

A decision tree classifier was trained on the processed training set using k-fold cross-validation to evaluate performance, improve robustness, and prevent overfitting. Model outputs were interpreted using term importance rankings and performance metrics, linking key terms such as “beauty” and “beautifully” with positive sentiment, and “nothing,” “disappoint,” and “remove” with negative sentiment.

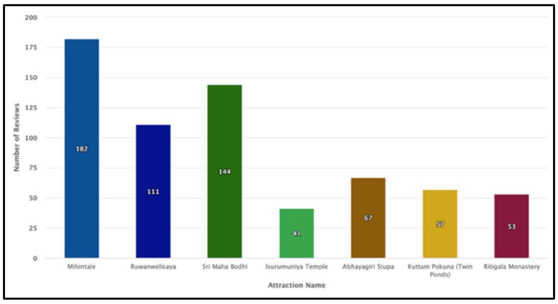
This methodology integrates systematic text preprocessing, dimensionality reduction, sentiment discretization, upsampling for class balance, and rigorous model evaluation to provide a structured, data-driven framework for analyzing tourist sentiment and supporting heritage tourism management strategies in Anuradhapura.

## 4. RESULTS AND DISCUSSION

Tourist perception of cultural and environmental values has been widely examined in recent studies, with particular emphasis on emotional engagement. This study examines the application of text mining and machine learning-based sentiment classification to analyse overall sentiment levels using tourist reviews of heritage sites in Anuradhapura, Sri Lanka, aiming to produce data-driven insights to support tourism planning.

The chart shown in figure 12 indicates that Mihinthale received the highest number of reviews (182), followed by Sri Maha Bodhi (144), while Isurumuniya Temple had the fewest reviews (41).

**Figure 12: Number of User Reviews of Each Location**




*Source: TripAdvisor and Google Reviews (2018–2025)*

### 4.1. Model Performance

The sentiment classification model, based on a decision tree algorithm, achieved an overall accuracy of 80.85% with a kappa of 0.634, indicating substantial agreement beyond chance. Figure 13 below includes the statistics of the confusion matrix.

**Figure 13: Confusion Matrix of Sentiment Classification**

|  | true Positive | true Neutral | true Negative | class precision |
|---|---|---|---|---|
| pred. Positive | 301 | 25 | 18 | 87.50% |
| pred. Neutral | 1 | 0 | 0 | 0.00% |
| pred. Negative | 105 | 14 | 389 | 76.57% |
| class recall | 73.96% | 0.00% | 95.58% |  |

*Source: Author (2025)*

It performed strongly in identifying positive (301 correct) and negative (389 correct) reviews, while neutral reviews were rarely captured. Weighted mean recall (56.5%) and precision (57.5%) were moderate, reflecting challenges in distinguishing minority classes. Overall, the model effectively captured the dominant positive and negative sentiment patterns in tourist reviews.

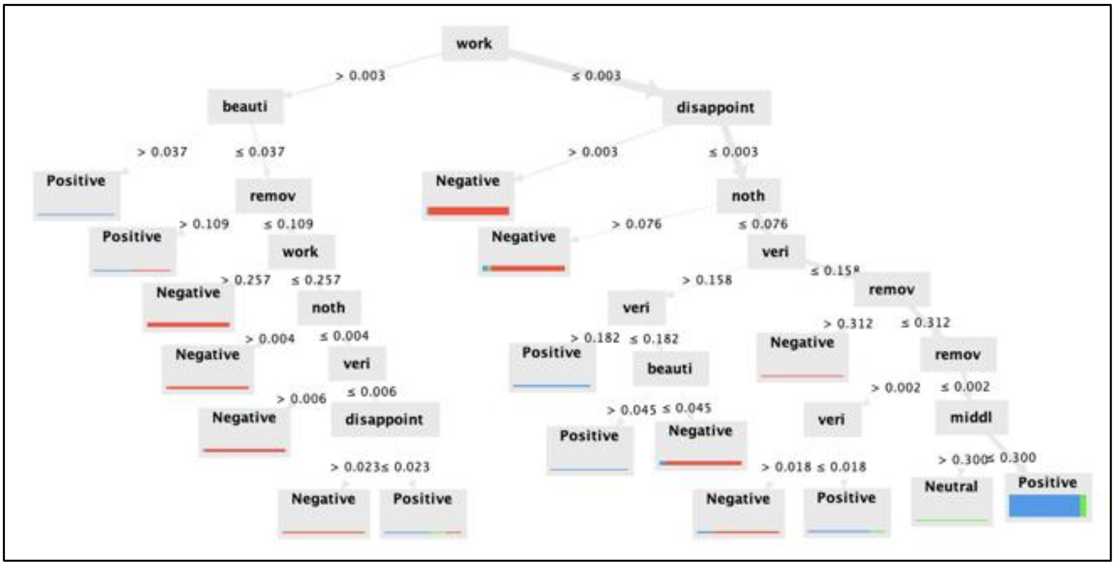
Compared with prior tourism sentiment studies that rely on deep learning or lexicon-based approaches (e.g., Viñan-Ludeña & de Campos, 2022; Gandy et al., 2024), the decision tree model used in this study demonstrates that interpretable machine learning techniques can achieve robust performance while maintaining transparency. This is particularly important in heritage tourism contexts, where explainability supports managerial trust and policy relevance. The limited detection of neutral sentiment also aligns with previous findings that neutral affect is often linguistically subtle and underrepresented in user-generated tourism content (Singgalen, 2023).

### 4.2. Sentiment Classification

The sentiment analysis of tourist reviews from Anuradhapura's heritage sites, conducted using a text mining approach with a decision tree-based classification model, revealed clear patterns in how specific words and phrases influence positive or negative evaluations. Consistent with prior research in heritage tourism sentiment analysis (Singgalen, 2023), key terms emerged as strong predictors of visitor sentiment. The output is shown in figure 14.

Words associated with aesthetics and admiration, such as "beauty" and "beautifully," were closely linked to positive sentiment. For instance, one visitor noted, "*The grounds are beautifully maintained, and there's a sacred atmosphere that invites you to pause and reflect*," highlighting how both visual appeal and emotional ambiance contributed to a highly favourable assessment. Another review emphasized architectural and scenic value, describing the brick stupa as "*a real Buddhist reliquary work… a beautiful walk in the park with some remains*," illustrating how

**Figure 14: Decision Tree of Sentiment Classification**




*Source: Author (2025)*

references to architectural achievement and the surrounding environment reinforce positive perceptions. The results of this study are consistent with the findings of Yuan et al. (2025), who reported that positive visitor sentiments are strongly associated with aesthetic appeal and cultural significance of heritage sites.

These findings reinforce prior research emphasizing the role of aesthetic appreciation and emotional engagement in heritage tourism satisfaction. Studies on World Heritage destinations consistently report that visual grandeur, symbolic meaning, and spiritual ambiance strongly shape visitor evaluations (Medina-Viruel et al., 2019; Yuan et al., 2025). Further, the strong association between aesthetic descriptors and positive sentiment suggests that visitors interpret heritage value not solely through historical knowledge but through immediate sensory and emotional responses. This supports the experiential tourism perspective, which argues that affective reactions often precede cognitive evaluation in culturally significant settings (Xiang et al., 2015). The prominence of such descriptors in this study suggests that affective responses play a dominant role in shaping international tourists' perceptions of Anuradhapura's heritage landscape.

Negative sentiment was most strongly associated with terms reflecting unmet expectations or dissatisfaction. The word "disappoint" emerged as a major indicator, appearing in reviews such as, "*In my opinion, in general, the whole site is pretty disappointing. As the country has better attractions, I would visit Anuradhapura only in case you have some spare time*," and was linked to 130 negative cases in the decision tree, highlighting its robust predictive power. Similarly, "nothing" appeared in reviews expressing underwhelming experiences, including, "*We came here and it was just another stupa! Nothing special to make your way here if you don't have time… There is nothing outstanding,*" with 72 negative cases captured in the tree. The term "work" was another significant predictor of negative sentiment, appearing in 63 cases, reflecting visitors' perception of effort or inconvenience in exploring certain sites. Secondary terms such as "remove" and "very" also appeared in the tree. "Remove" was associated with operational discomfort, particularly the need to remove shoes or hats or parasols and was linked to 13 negative cases and several positive ones depending on context, as reflected in the reviews: "*No shoes allowed in the complex, as usual, but this gets extremely uncomfortable when the weather is hot," and "We arrived at the hottest part of the day to be told that we could not wear hats or even parasols… Very unpleasant experience. Avoid."* These patterns mirror findings in other heritage tourism studies, such as Viñan-Ludeña and de Campos (2022), who identified operational inefficiencies and service delays as key drivers of visitor dissatisfaction, and Yuan et al. (2025), who found that site management and environmental factors, including safety and comfort, were major contributors to negative visitor perceptions.

The strong association between negative sentiment and operational or experiential constraints mirrors findings from earlier tourism studies, which identify site management, accessibility, and visitor comfort as key sources of dissatisfaction (Viñan-Ludeña & de Campos, 2022; Yuan et al., 2025). However, in contrast to urban heritage settings, the Anuradhapura context highlights climate-related discomfort and religious protocol (e.g., shoe and hat removal) as particularly influential, suggesting that destination-specific cultural practices must be carefully balanced with international visitor expectations.

The term "Very," functioning primarily as an intensifier, contributed to sentiment classification in 35 negative cases and 16 positive cases when combined with other terms, demonstrating that intensifiers can subtly modulate perceived sentiment. The term “middl” captured neutral sentiment in a small subset of cases (32), indicating reviews that were neither strongly positive nor negative. However, the identification of intensifiers such as “very” as sentiment-modulating terms aligns with NLP literature emphasizing the contextual sensitivity of affective language. This finding underscores the importance of interpretable models, such as decision trees, which allow nuanced linguistic patterns to be examined directly, an advantage often obscured in black-box sentiment classifiers.

Overall, the findings emphasize the multidimensional nature of tourist experiences at heritage sites. Positive reviews were primarily shaped by aesthetic appeal, architectural significance, and spiritual engagement, whereas negative reviews reflected unmet expectations, operational challenges, and underwhelming site conditions. Decision tree-based sentiment classification effectively captured these patterns, providing nuanced insights into how specific textual cues signal broader visitor experiences. Such insights can guide heritage tourism managers in enhancing visitor satisfaction by emphasizing admired features while addressing sources of dissatisfaction.

Taken together, these results demonstrate how machine learning-based sentiment analysis can bridge empirical evidence and tourism theory by translating large-scale visitor narratives into interpretable patterns of perception. By situating sentiment outcomes within existing literature on emotional engagement, destination experience, and heritage management, the study moves beyond descriptive analytics to offer theoretically grounded insights for both researchers and practitioners.

### 4.3. Methodological Implications

This study makes several important methodological contributions to tourism analytics and heritage research. First, it demonstrates the effectiveness of integrating large-scale user-generated content from multiple online platforms (Google Reviews and TripAdvisor) as an alternative to traditional survey-based approaches, enabling the analysis of authentic, unsolicited, and temporally rich visitor narratives. This approach addresses common limitations of conventional tourism research, such as small sample sizes, recall bias, and static representations of visitor experience.

Second, the study proposes a structured and replicable text mining pipeline that combines systematic preprocessing, feature selection through dimensionality reduction, sentiment discretization based on star ratings, and class imbalance correction using SMOTE. This methodological configuration enhances model robustness and provides a practical framework for handling noisy, imbalanced tourism review data, an issue frequently encountered but rarely addressed explicitly in heritage tourism studies.

Third, the use of a decision tree classifier offers a transparent and interpretable machine learning approach, allowing researchers and practitioners to directly link specific words and phrases to sentiment outcomes. Unlike black-box models, this interpretability is particularly valuable in heritage and policy-oriented contexts, where explainability is essential for informed decision-making, stakeholder communication, and operational interventions.

Finally, the study demonstrates the scalability and transferability of machine learning-based sentiment analysis for heritage destination management. The methodological framework can be readily adapted to other cultural destinations, languages, and tourism contexts, supporting real-time monitoring of visitor sentiment and evidence-based planning. As such, the study contributes not only empirical insights but also a methodological template for future tourism, heritage, and destination management research.

### 4.4. Managerial and Policy Level Implications

This study demonstrates the policy relevance and transformative potential of machine learning–based sentiment classification in tourism analytics, particularly for heritage-rich destinations such as Anuradhapura. By systematically analysing online international tourist reviews, the study shows how user-generated content can be leveraged as a continuous, data-driven feedback mechanism to inform tourism policy formulation, heritage conservation strategies, and destination governance. The findings highlight the value of sentiment analytics as a complementary evidence base to traditional surveys and administrative data, enabling policymakers to capture real-time visitor perceptions at scale.

At the policy level, the results emphasize the importance of prioritising visitor experience as a core component of heritage tourism planning. Positive sentiment associated with aesthetic quality, architectural heritage, and spiritual ambiance suggests that policies should continue to support conservation-led development that preserves the visual, cultural, and symbolic integrity of heritage sites. At the same time, negative sentiment linked to operational discomfort, unmet expectations, and site management challenges signals the need for targeted policy interventions aimed at improving basic visitor infrastructure and service standards.

Tourism and heritage authorities can translate these insights into concrete policy actions by allocating resources toward visitor comfort and accessibility. Policies supporting the provision of shaded rest areas, water points, seating along long or steep pathways, and practical solutions such as protective mats at sites requiring shoe or hat removal can significantly enhance visitor satisfaction, particularly in hot climatic conditions. In addition, policies should promote improved interpretive and educational infrastructure, including multilingual signage, explanatory boards, and the development of digital interpretation tools or mobile applications that use audio-visual storytelling to communicate historical and cultural significance more effectively to international visitors.

The findings also have implications for pricing and revenue policies. Clear communication of entry fees and alignment of pricing with the perceived quality of the visitor experience can help manage expectations and reduce dissatisfaction. Revenue generated through entry fees should be transparently reinvested in site maintenance, conservation activities, and visitor services to sustain long-term destination quality and credibility.

Finally, the study highlights the need for data-driven governance mechanisms in tourism policy. The integration of sentiment dashboards into tourism management systems would allow policymakers to monitor visitor perceptions in near real time, identify emerging issues, and evaluate the effectiveness of policy interventions and operational changes. Such systems can support adaptive policymaking, guide strategic marketing decisions, and improve coordination between tourism authorities, heritage managers, and local governments. Given its scalability and cost-effectiveness, this approach is particularly well suited for heritage destinations in developing-country contexts, enabling more responsive, evidence-based, and sustainable tourism policy development.

### 4.5. Limitations and Future Research Directions

This study has several limitations. First, the analysis relies primarily on online reviews and user-generated content, which may be biased toward more vocal or digitally active visitors, potentially underrepresenting other visitor segments.

Second, the research focuses on specific heritage sites, limiting the generalizability of the findings to other locations or types of tourism properties. Third, the sentiment analysis is based on textual data and may not fully capture nuanced visitor experiences, such as emotions or contextual factors influencing satisfaction. Further, contextual ambiguities such as sarcasm or mixed sentiments, may also complicate accurate classification. Finally, external factors like seasonal variations, special events, or temporary disruptions were not extensively controlled for, which may influence visitor perceptions.

Future studies could incorporate mixed methods, such as surveys, interviews, or observational studies, and complement textual analysis and provide a deeper understanding of visitor experiences. Research could also explore the impact of targeted interventions, such as infrastructure improvements or digital interpretation tools, on visitor satisfaction over time. Finally, future studies should also explore aspect-based sentiment analysis to capture opinions on specific attributes like cleanliness and accessibility, integrate spatio-temporal data to visualize sentiment trends over time and location, profile tourist personas by combining sentiment with demographic data, and design hybrid recommender systems that incorporate sentiment insights for personalized travel planning.

## 5. CONCLUSION

This study demonstrates the value of applying text mining and decision tree-based sentiment classification to analyze tourist reviews of Anuradhapura's heritage sites. The findings reveal clear patterns in visitor perceptions, highlighting both strengths, such as aesthetic appeal and cultural significance, and areas needing improvement, including operational challenges and unmet expectations. By providing actionable, data-driven insights, this approach enables heritage managers, policymakers, and marketers to enhance visitor experiences, preserve cultural assets, and implement targeted interventions. The methodology offers a scalable, adaptable framework for heritage tourism analytics in Sri Lanka and beyond, supporting evidence-based, sustainable tourism development. In addition, beyond its empirical findings, this study contributes to the tourism analytics literature by demonstrating how interpretable machine learning models can be systematically applied to heritage tourism contexts, offering both methodological transparency and theoretically informed insights into visitor perception.

## 6. ACKNOWLEDGEMENT

The author gratefully acknowledges Professor Rens Scheepers, Department of Information Systems and Business Analytics, School of Business and Law, Deakin University, Australia, for his guidance and insights, which provided valuable understanding of text mining and sentiment classification applied in this study.

## 7. REFERENCE

Analytics Vidhya. (2025). Retrieved from https://www.linkedin.com/pulse/what-sentiment-analysis-claudiu-clement/

Cambria, E., White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine, 9* (2), 48-57.

Diabolocom. (2023). Natural Language Processing (NLP): a complete guide. Retrieved from https://www.diabolocom.com/blog/natural-language-processing-nlp/

Gulati, S. (2022). Tapping public sentiments on Twitter for tourism insights: A study of famous Indian heritage sites. *International Hospitality Review, 36*(2), 244–257. https://doi.org/10.1108/IHR-03-2021-0021

Kok, J.N., Boers, E.J. & Kosters, W.A. (2009) Artificial intelligence: Definition, trends, techniques, and cases. Artificial Intelligence 1: 270–299

Mutalib, S., Razali, A. H., Kamarudin, S. N. K., Halim, S. A., & Abdul-Rahman, S. (2021). Prediction of tourist visit in Taman Negara Pahang, Malaysia using regression models. *International Journal of Advanced Computer Science and Applications, 12*(12), 746–754. https://doi.org/10.14569/IJACSA.2021.0121292

Punchihewa, S. (2021). *Recognition of Mihintale as a World Heritage Site is long overdue*. [Photograph]. *The Island*.

Singgalen, Y. A. (2023). Analisis Sentimen Top 10 Traveler Ranked Hotel Di Kota Makassar Menggunakan Algoritma Decision Tree Dan Support Vector Machine. *KLIK: Kajian Ilmiah Informatika dan Komputer, 4*(1), 323–332.

Surugiu, C., Surugiu, M.-R., & Grădinaru, C. (2023). Targeting creativity through sentiment analysis: A survey on Bucharest city tourism. *SAGE Open, 13*(2), 1–17. https://doi.org/10.1177/21582440231167346

Travel Setu. (2025). *Jetavanaramaya Tourism Guide*. [Photograph]. Retrieved from https://travelsetu.com/guide/jetavanaramaya-tourism

Tripadvisor. (2025). *Abhayagiri Dagaba, Anuradhapura*. [Photograph].

Viñan-Ludeña, M. S., & de Campos, L. M. (2022). Discovering a tourism destination with social media data: BERT-based sentiment analysis. *Journal of Hospitality and Tourism Technology, 13*(5), 907–921. https://doi.org/10.1108/JHTT-09-2021-0259

Wonders of Ceylon. (2024). *Isurumuniya Temple, Anuradhapura*. [Photograph]. Retrieved from https://www.wondersofceylon.com/isurumuniya-temple/

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

Yusof, N. N., Mohamed, A., & Abdul-Rahman, S. (2019). Context enrichment model-based framework for sentiment analysis. In Communications in Computer and Information Science (Vol. 1100, pp. 325–335). Springer. https://doi.org/10.1007/978-981-15-0399-3_26

Yuan, H., Ke, R., & Xie, X. (2025). Sentiment analysis of visitor perceptions on architectural heritage: A case study of Phoenix Ancient Town for sustainable conservation and development. *Journal of Asian Architecture and Building Engineering*. Advance online publication. https://doi.org/10.1080/13467581.2025.2540079

Zhou, L., Song, Y., Ji, W., & Wei, H. (2022). Machine learning for combustion. *Energy and AI, 7,* 100128. https://doi.org/10.1016/j.egyai.2021.100128