Destination Image Through Text Analysis of Uk Based International News Media: A Study on Sri Lanka Tourism

Received 15 December 2025
Revised 27 January 2026
Revised 19 March 2026
Accepted 4 July 2026
Published (Online) 15 August 2026

Chirasti Amarasinghe
Undergraduate, University of Moratuwa
Kattuebedda Sri Lanka
chirasthi.sy@gmail.com

Indra Mahakalanda
Department of Decision Sciences, Faculty of Business,
University of Moratuwa, Kattuebedda Sri Lanka
indram@uom

Abstract

Understanding how international news media shapes a country's tourism sector is challenging because of the wide variety of media sources, perspectives and understanding. Sentiment analysis plays a crucial role, especially in tourism, where understanding the public's view-point is key for effective marketing strategies and destination management. This study presents a novel approach to evaluating how Sri Lanka's tourism industry is portrayed in international news in the United Kingdom. By applying sentiment analysis models and ensuring accuracy through careful data collection, cleaning, and preparation, the research identifies clear patterns linking positive media sentiment with increased tourist arrivals to the island. Additionally, examining media coverage over time helps reveal shifts in sentiment, offering deeper insight into how Sri Lanka's image as a travel destination has developed globally, particularly in the English-speaking countries. Overall, this comprehensive analysis provides valuable guidance for stakeholders to enhance and manage the country's tourism reputation through focused promotional efforts.

Keywords: Sentiment analysis; Text analytics; Destination image; Sri Lanka; Content analysis; Tourism.

SLJMS

Sri Lanka Journal of Management Studies
Vol. 8 – Issue I, 2026
pp, 23 – 38
ISSN (Print): 2682-7298
ISSN (Online): 2792-1093
https://doi.org/10.4038/sljms.v8i1.213

Introduction

The role of media in shaping the Destination Image (DI) of a country is increasingly becoming significant in the digital age. Tourists often form their perceptions based on what media portrays before making travel-related decisions due to the widespread access to information. It influences the pre-visit destination image and continues to shape perceptions even after the visit, reflecting the ongoing image held by tourists (Gartner, 1993; Govers et al., 2007).

By disseminating these contents, one can gain valuable insights into the perceptions and sentiments associated with a particular location. It can aid in identifying many insights such as which media outlets contribute to a positive image, identifying seasonal trends and key markets. This would enable policymakers to formulate strategically targeted promotional campaigns that align with tourists.

While prior research has emphasized the role of social media, online platforms, and word-of-mouth communication in shaping DI (Ghazali & Cai, 2013), the influence of international news media remains comparatively underexplored. News coverage, particularly of crises, political conditions, or natural disasters plays a critical role in forming perceptions of destinations by reflecting real-time events and shaping global narratives (Stepchenkova & Eales, 2011). More than the promotional content, news reporting often highlights risk, safety, and socio-political issues of destinations. This makes news reporting a valuable source of information for potential tourists. Despite this importance, there is limited empirical evidence on how DI evolves when framed through international news sentiment. To address this gap, this study analyses destination image by how it is portrayed in news media appeared in the United Kingdom (UK) to gain insights that would be useful for the uplifting of the tourism sector. For this purpose, Sri Lanka was chosen as the case destination.

Situated in the Indian Ocean, Sri Lanka possesses a rich culture along with many scenic locations, making it a popular destination for travelers. This has contributed significantly to the tourism industry making it the third-largest export earner (Mitchell, 2022). In 2019, 1.9 million tourists visited Sri Lanka generating $3.6 billion in foreign income. This has helped in sustaining the economy, creating jobs and business opportunities for the country. Much of it can be attributed to the Sri Lankan destination image.

Despite the tourism industry's significant accomplishments, Sri Lanka has recently faced persistent difficulties in managing its destination image and promoting tourism. In recent years, the county has grappled with negative media coverage surrounding incidents such as the COVID-19 pandemic and the protests. Thus, negative perception of Sri Lanka has led to reduction of tourism (Arachchi & Gnanapala, 2020).

The gap between Sri Lanka's significant tourism potential and the effects of recurring crises makes it an interesting case for studying how media-reported external shocks influence destination image over time. Analyzing the tone and sentiment of news coverage about Sri Lanka offers valuable insights into how it is perceived by international audiences, which can support the development of more targeted advertising strategies, and improve Sri Lanka's viability in tourism hence driving growth in the tourism sector.

Although Sri Lanka draws tourists from a wide range of markets, this study focuses specifically on news media in the United Kingdom. The UK consistently ranks among the country's top sources of inbound tourism, accounting for around 12% of total arrivals (Sri Lanka Tourism Development Authority, 2025). In addition, UK media holds significant global influence, making it a suitable lens through which to examine how international audiences perceive Sri Lanka. Accordingly, this study analyzes shifts in Sri Lanka's image by applying text analytics and sentiment analysis to UK news coverage.

The main objective is to develop a text analytics model to analyze designation image in tourism of Sri Lanka trough the UK based international media. This study makes two main contributions. First, by incorporating the UK based international news media as a central factor in image formation, it extends the theoretical scope of destination image (DI) literature. Second, it contributes to the expanding research on how destination images evolve dynamically in response to external shocks rather than remaining fixed. From a practical standpoint, the findings offer valuable insights for policymakers and tourism marketers, helping them craft communication and branding strategies that counter negative portrayals while strengthening positive narratives. We have selected UK based international news media to avoid computational difficulties of natural language processing. The remainder of the paper is structured as follows. Section 2 provides an overview of the literature that discusses previous studies of destination images. Section 3 describes the methodology followed for the conduct of the study while Section 4 presents our analysis and discusses the results obtained. The paper concludes with Section 5 discussing the practical implications and limitations of the study and avenues for future research.

Literature Review

DI has been discussed extensively in tourism research due to its critical role in influencing travelers' perceptions and decision-making. Crompton's seminal work (1979) introduced this definition of DI that opened a new avenue of tourism research. It is a highly abstract and complex concept that ranges from ideas, impressions to beliefs that a person holds of a destination (Wang et al., 2023). This view is further enhanced by Gartner (1993) who defined DI as being composed of three components of cognitive image, affective image, and conative image. Zhang et al. (2014) concluded that the destination image has a significant impact on tourist loyalty. Chia et al. (2021) suggested that the political and cultural image of a country has a positive relationship with satisfaction and that it leads to tourist retention and destination loyalty.

In the analysis of DI, both qualitative and quantitative methods have been employed in previous studies. Qualitative methods such as interviews, focus groups, and content analysis provide in-depth insights into tourists' perceptions (Dwyer et al., 2012). Tseng et al. (2015) examined how blogs helped shape travelers' perception of China. This was achieved by performing a qualitative analysis of the content in blog articles. Februadi et al. (2019) explored the destination image of Bandung utilizing interview data collected from domestic tourists. They were processed using descriptive statistical methods and content analysis.

Conversely, quantitative methods use statistical techniques to quantify the relationship between DI and tourism. These methods provide generalizable findings due to the data collected from large samples. De Nisco et al. (2015) surveyed 542 tourists and utilized statistical techniques to conclude that DI influences tourism satisfaction on post-behavioral intentions. While qualitative methods excel in capturing the depth and complexity of destination image, quantitative methods offer generalizable results.

Natural Language Processing (NLP) techniques have become increasingly popular in destination image research as they allow the analysis of large volumes of text data from various sources. Text classification, topic modeling, and sentiment analysis are widely used in this analysis of textual data. Guerrero-Rodriguez et al. (2021) carried out an analysis of online travel reviews related to Guanajuato using NLP approaches. They were used to extract the themes and topics prevalent within the reviews.

Sentiment analysis techniques are widely used in NLP to assess the emotional tone of text data related to destination image. Most literature utilizes lexicon-based and machine learning based approaches when analyzing sentiment. Lexicon-based techniques rely on predefined dictionaries to determine the sentiment of the text (Liu, 2022). While they are easy to implement, they struggle with context dependent words and meanings. On the other hand, machine learning based approaches involve training algorithms on datasets, to recognize sentiment patterns in text. They have shown higher accuracy and adaptability in comparison. Jiang et al. (2021) utilized machine learning techniques to assess the destination image of Hong Kong using online reviews from tourism websites. Liu et al. (2020) have also explored the change in the destination image of Macau using text mining techniques. This has revealed that Chinese tourists are more sensitive to new attractions than international tourists are.

Although machine learning techniques have been popularly used in recent destination image literature, transformer-based models such as BERT and GPT have not yet been widely adopted. These models reflect a significant advancement in NLP as they are better capable of capturing contextual information and meanings in comparison to traditional models (Devlin et al., 2018). Integrating transformer models into destination image research could lead to a more comprehensive analysis that traditional models may not capture.

Furthermore, recent literature shows that sentiment analysis has predominantly relied on tourist reviews. These are valuable since they provide personal experiences of travelers. However, news articles provide a strong alternative data source. They are able to provide a comprehensive perspective as it covers a variety of subjects. They are also considered credible and are used as an important public information source (Chen et al., 2020). Despite this potential for improved analysis, news articles have not been utilized much in literature.

There are noticeable gaps in the vast amount of literature that examines the complexities of the tourism sector. First, previous research has mostly relied on tourist-generated content such as reviews or social media posts, while neglecting international news as a reliable and agenda-setting source that is capable of influencing tourist attitudes. Second, limited attention has been given to countries characterized by socio-political volatility where media narratives play a decisive role in tourism recovery. Third, while NLP and sentiment analysis have advanced DI research potential, few studies have applied state-of-the-art transformer models to analyze the nuanced and evolving tone of news content. This study seeks to address these gaps by integrating text analysis with destination image theory to assess how Sri Lanka's image has evolved in UK media coverage. This is especially significant now as digital media reshapes public opinion.

Methodology

The methodology, as indicated in Figure 1, forms the foundational basis upon which this study is built, providing a theoretical lens to address the research aim of analyzing the international perception of Sri Lanka. The methodology begins with the data collection that would be preprocessed using text analytics. Afterward, the preprocessed data would be used to develop sentiment models for further analysis.

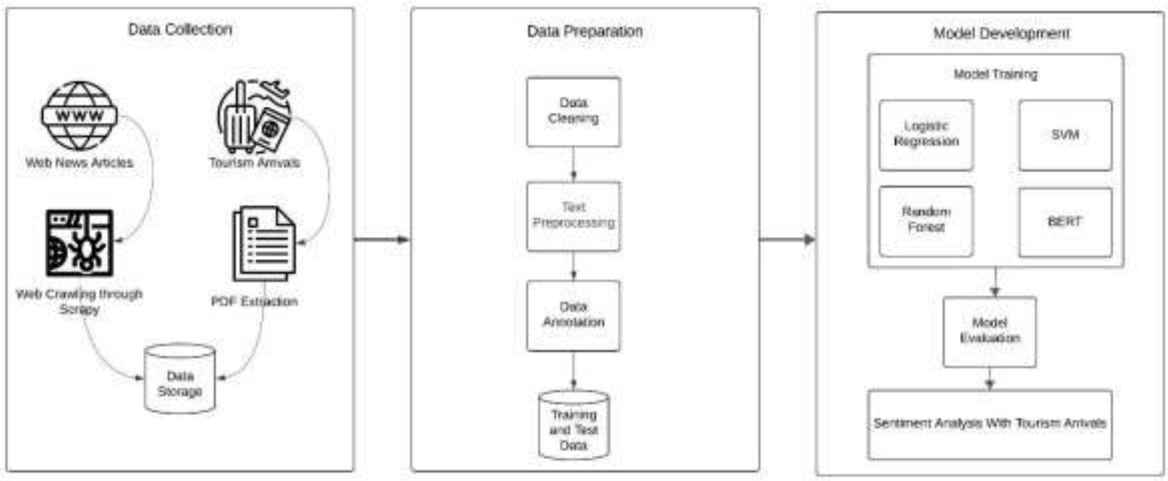

Figure 1: High Level View of Solution Approach

Data Collection and Preprocessing

Accordingly, the study is initiated by the collection of necessary data. Official Sri Lankan tourism statistics and monthly tourist arrivals were extracted from annual tourism reports published by the Sri Lankan Tourism Development Authority. The main source of data for which the sentiment analysis is to be carried on is extracted from well-known news websites in the UK. A sample of daily news articles from BBC, Guardian, and the Times that mention the keyword "Sri Lanka" within the period of 2014 to 2023 were selected minimizing bias during filtering. These outlets were selected due to their global reach and influence in shaping international perception. Dataset could be considered as bias, if the results are generalized to audiences such as non-English speaking and non-frequent mainstream media. As well-established media organizations in the UK, they have built a solid audience that reaches more than 40 million citizens (Nielsen et al., 2025). This aids in creating a representative sample of data for the sentiment analysis.

The extraction process was automated with the use of web scraping via the Python library, Scrapy (Zyte, 2023). Scrapy is a tool that creates spiders that can crawl and extract large amounts of data. Custom scraping scripts (spiders) were developed to crawl through the built-in search functionality of each news website using the keyword "Sri Lanka" systematically navigating links where available.

After the articles have been extracted, the data were preprocessed before analysis and model building. Duplicate articles arising from repeat indexing across search pages were identified and removed. In addition, any entries that did not contain substantive article content were excluded from further analysis. This was conducted with the use of several Natural Language Processing (NLP) techniques such as tokenization, and lemmatization using the Spacy library (SpaCy, 2024). First, all text was converted to lowercase. Thereafter, general stop words such as "the, is, and" are removed. Since these high frequency terms add no value to the task, these words were removed. Additionally, all irrelevant phrases and characters such as hyperlinks, HTML tags, hashtags, and metadata were removed using the Regular Expressions library (Van Rossum, 2020).

Model Development: Sentiment Analysis

Sentiment analysis is an NLP technique that is capable of expressing the sentiment present in text (Ain et al., 2017). This aids in gauging the emotions expressed by the writer or emotions felt by the reader. This technique is particularly useful in this study as we analyze the destination image of Sri Lanka held by tourists when reading news about Sri Lanka.

However, before applying sentiment analysis, it was necessary to treat the class-imbalanced nature of the dataset as the unequal distribution of sentiment classes could result in a biased model performance with poor performance in the minority sentiment (Lango, 2019). From the many ways of treating class imbalance class weights, random oversampling, and random under sampling were tested for each of the models built for sentiment analysis. Higher weights are assigned to minority classes and vice versa in the class imbalance method. Random oversampling is where more data is sampled from the minority class while random under sampling is where lesser data is sampled from the majority class. Random oversampling was selected as it produced more consistent performance across evaluation metrics. All words of "Sri Lanka" were removed at the time of model development.

After treating the class imbalance, sentiment analysis was performed using several algorithms. For machine learning algorithms Logistic Regression, Support Vector Machine (SVM), Random Forest, and transformers were used. These models were selected to investigate different approaches from conventional algorithms to advanced techniques.

The Logistic Regression model assigns an observation into a particular class based on the probability of it belonging to that particular class (Indra et al., 2016). It uses input features to calculate the probability that will be used to assign the class label. SVM is a powerful classification algorithm that identifies the optimal hyperplane that can effectively classify observations into different classes. It performs well in text classification tasks because it can handle bother linear and non-linear decision boundaries (Fabrice & Pavel, 2006).

On the other hand, Random Forest models are ensemble learning models that use multiple decision trees (Karthika et al., 2019) in its classification approach. The model's ability to not overfit makes it a suitable algorithm for text categorization. Transformer models employ a multi-head attention mechanism that simultaneously assesses text input from multiple angles. This allows it to capture different types of relationships within it. They are capable of learning context and tracking semantic dependencies in text. For this investigation, the Bidirectional Encoder representation from Transformers (BERT) base-uncased transformer was employed (Devlin et al., 2018).

Model Evaluation

Model evaluation is required to ensure that sentiment was assessed appropriately. This guarantees that further analyses conducted are reliable and accurate. All of the models created for this purpose were tested on several performance metrics that are summarized in Table 2.

Table 1: Tested Performance Metrics

Accuracy	The proportion of correct predictions out of total predictions	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
Precision	The proportion of true positive predictions out of all positive predictions	$Precision = \frac{TP}{TP + FP}$
Recall	The proportion of true positive predictions out of all actual positive instances	$Recall = \frac{TP}{TP + FN}$
F1 Score (weighted)	The harmonic means of precision and recall	$F1\ Score = \frac{2(Recall \times Precision}{Recall + Precision}$
Area under the Receiver Operating Characteristic (ROC) curve	A measure of the model's ability to distinguish between classes	$AUC = \int_0^1 TPR(FPR)\, d(FPR)$

Source: Survey Data

Since the dataset is imbalanced, using accuracy alone presents a biased evaluation. Therefore, testing the model on several measures is necessary to ensure a good fit. The weighted F1 score is used for the final evaluation as it provides more insight into the functionality of the model by calculating an average of precision and recall. The results of applying this methodology, including the performance of the models and their corresponding evaluations, are presented in the next section.

Results

This section investigates the detailed results and insights obtained from the methodology including the models trained for sentiment analysis and tourism arrivals.

The content of 1607 news articles were scraped using the Scrapy framework. Along with the content, meta data such as the date published, the headline, and the source were also collected. The breakdown of the selected sources is shown in Table 1. The scraped data was then processed, cleaned, and stored using Scrapy's pipeline tools.

Table 2: Data Set

News Website	Number of Scraped Articles
BBC	329
Times	552
Guardian	726

Source: Survey Data

The articles were then preprocessed using tokenization and lemmatization techniques via the Spacy library (SpaCy, 2024).

The restructured articles were subsequently annotated by hand to support the sentiment analysis task. The annotation process was carried out independently by multiple annotators following predefined sentiment classification guidelines. Any discrepancies were resolved through disucssion to produce a final agreed-upon label. This resulted in 483 positive articles and 1124 negative articles.

Exploratory Data Analysis was conducted initially to understand the nature of the data. Figure 2 presents the word clouds created from the collected articles for each sentiment category. In the negative sentiment cloud, prominent words such as "government," "asylum seeker," "police," and "prime minister" point to themes of political instability and recent government actions in Sri Lanka. These terms highlight public concerns related to governance, human rights, and social unrest.

In contrast, the positive sentiment word cloud features words like "people," "day," "family," "beach," and "hotel," which are strongly linked to hospitality and enjoyable experiences. Additional terms such as "local," "visit," and "tour" emphasize Sri Lanka's attractiveness as a tourist destination, while frequently appearing words like "new" and "year" suggest possible growth and renewed opportunities in the tourism sector.

Figure 2: Word Cloud for each Sentiment

Sentiment Analysis

The sentiment analysis was performed by training models under several algorithms as mentioned previously. The labeled dataset was divided into training and testing sets based on time. Articles from the earlier period (2014 – 2021) were used for training while recent data was reserved for testing. This approach was adopted to preserve the chronological structure of the data. They were evaluated based on several metrics to ensure a better fit which are shown in Table 3. However, the weighted F1 Score was used as an unbiased measure to select the best model for sentiment analysis instead of accuracy, as there was a moderate class imbalance present. From the scores recorded, it is evident that the BERT classifier outperformed the rest of the model with an F1-score of 0.9112 followed by SVM with a score of 0.8731. The ROC curves indicated in Figure 3 show that the area under the curve is highest under the BERT model. The trained BERT Classifier model also provides the best results for other metrics indicating its superior ability to gauge news sentiment. Therefore, the BERT Classifier model is chosen for the analyses conducted ahead.

Table 3: Performance of Algorithms

Model	**Performance Metrics**			
	Accuracy	**F1 Score**	**Precision**	**Recall**
Logistic Regression	0.8711	0.8708	0.8381	0.92
SVM	0.8733	0.8731	0.8415	0.92
Random Forest	0.8043	0.7836	0.8269	0.4433
BERT Classifier	0.9111	0.9112	0.9149	0.9153

Source: Authors' Estimation

The sentiment model was then used to derive the sentiment scores for each article. The sentiment score ranges from zero to one, with zero indicating negative sentiment and one indicating positive sentiment. A linear sentiment score enables us to easily quantify the intensity and polarity of a text (Chiorrini et al., 2021). Figure 4 illustrates the sentiment scores over time, spanning from 2014 to early 2024. There is a significant volatility in perception Sri Lanka internationally.

The rising and falling pattern demonstrate how coverage has varied over time, where higher peaks indicate more positive portrayals and lower points represent phases of negative sentiment.

The years 2014 and 2016 exhibit notably high sentiment levels, aligning with positive developments in Sri Lanka. During this time, the country experienced economic growth following the civil war, which contributed to a rise in tourism and investment. In contrast, the decline observed in 2019 points to negative events that impacted the country's image, most likely linked to the Easter Sunday attacks (Kumarasiri et al., 2019).

After 2020, sentiment scores appear relatively stable, although they remain on the lower side. From 2021 onward, there is a steady upward trend in sentiment, reflecting a gradual recovery from earlier setbacks, including the effects of the COVID-19 pandemic and political instability. By 2023 and 2024, sentiment levels are both higher and more consistent, suggesting increasingly positive perceptions of Sri Lanka in international media. This improvement is likely driven by progress in tourism management, enhanced political governance, and overall economic recovery.

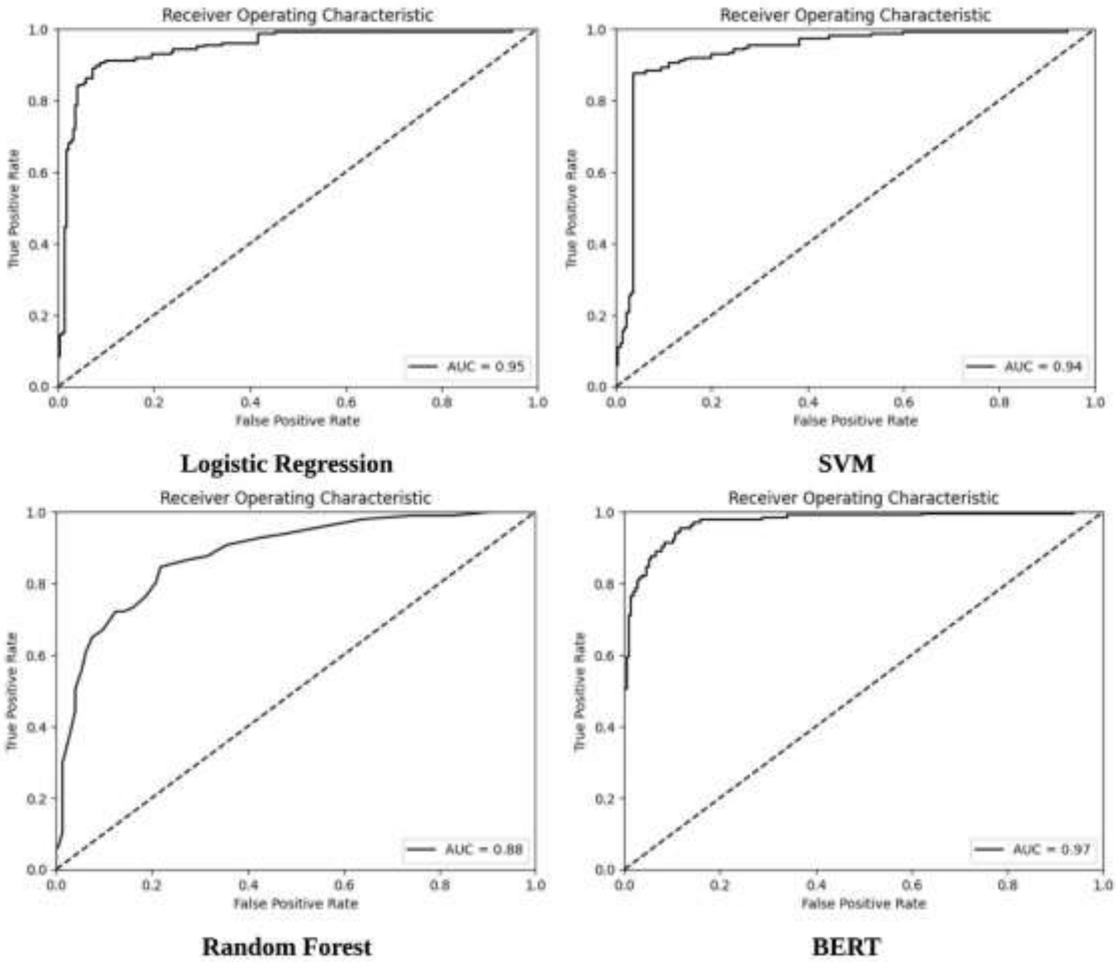

Figure 3: ROC Curves of Sentiment Analysis Models

Content Analysis

Sentiment scores were examined in relation to monthly tourist arrivals from the UK to Sri Lanka, as illustrated in Figure 5. Tourist numbers fluctuate significantly over time, with peaks often corresponding to earlier increases in positive sentiment and declines following periods of negative sentiment. Notably, higher arrival figures are seen during 2015–2016 and 2018–2019, which align with favorable international media coverage and effective tourism campaigns. A marked drop in arrivals occurs in early 2020, coinciding with the outbreak of the COVID-19 pandemic, which severely impacted global travel and tourism.

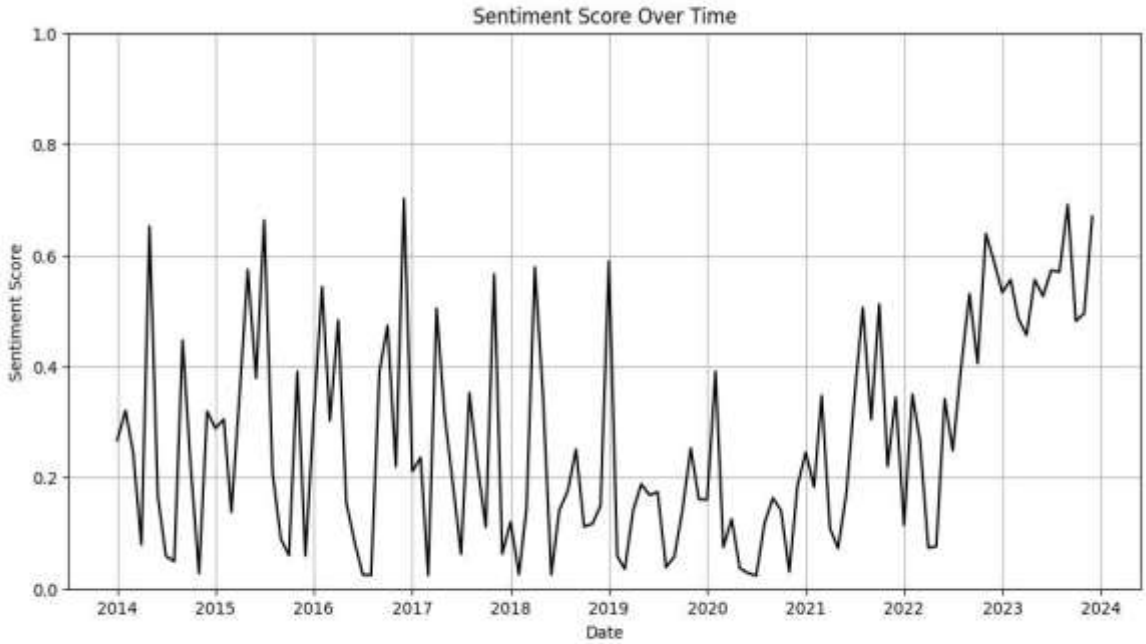

Figure 4: Change in Sentiment Score over Time

Overall, increases in sentiment scores reflecting more positive news coverage are generally followed by rises in tourist arrivals. In contrast, negative sentiment periods tend to be associated with reduced tourist numbers, due to the fact that unfavorable news discourages potential visitors.

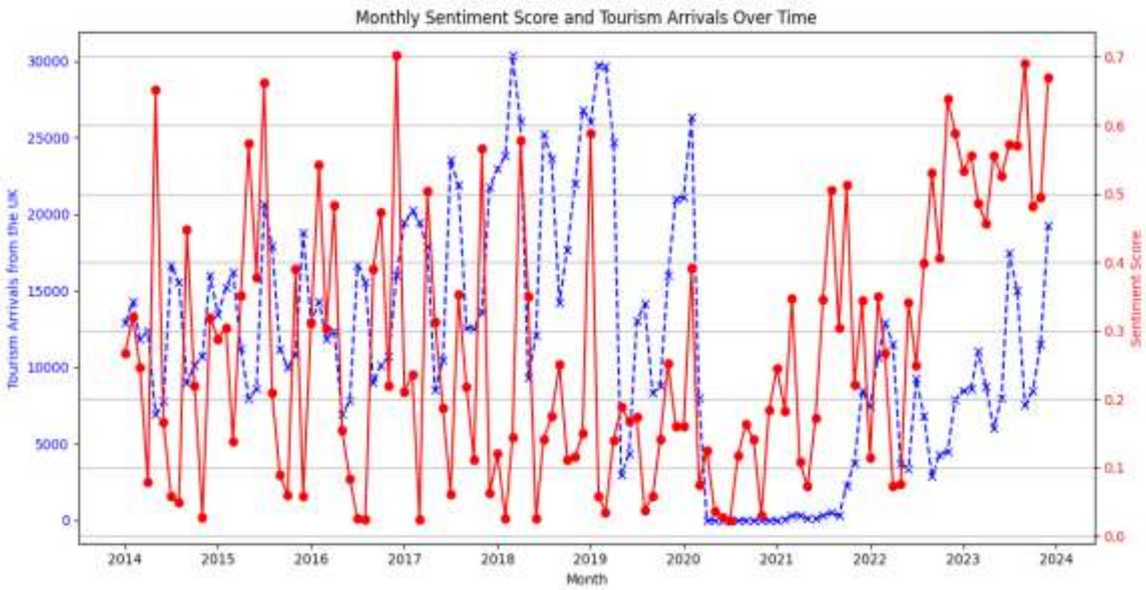

Figure 5: Relationship between Sentiment Score and Tourist Arrivals in Sri Lanka over Time

Key observations include the positive sentiment peak in early 2014 aligning with a high in tourist arrivals, possibly due to the post-war recovery phase where Sri Lanka was increasingly seen as a safe and attractive tourist destination. The significant dip in sentiment in April 2019 aligns with the Easter Sunday bombings, leading to a sharp decline in tourist arrivals due to security concerns. The impact of the COVID-19 pandemic is visible in sentiment, with tourist arrivals dropping sharply in early 2020 and gradually recovering as travel restrictions were lifted, and the situation stabilized.

It is important that the above observations be quantified to derive insights into the sentiment analysis. Therefore, the correlation was tested between lagged news sentiment and tourism arrivals from the UK. A secondary analysis was also conducted by deriving sentiment scores for various parts of the article such as the headline and first paragraph to understand the impact on tourism perception that is shown in Table 4. In order to test this relationship, the Spearman rank correlation coefficient was used (Bae & Lee, 2012). It is a non-parametric measure of the monotonic relationship between two ranked variables. This metric is better suited for the study as the relationship between sentiment and tourist arrivals is not linear in nature.

Table 4: Correlation Coefficients over Lagged Months

Lagged Number of Months	**Spearman Correlation Coefficient Values (* indicates significant coefficients)**		
	Full Article Content	**Headline**	**First Paragraph**
1 Month	0.189*	0.163	0.186
2 Months	0.123	0.141	0.135
3 Months	0.153	0.118	0.155
4 Months	0.141	0.136	0.135
5 Months	0.187*	0.165	0.173
6 Months	0.215*	0.211	0.187
7 Months	0.277*	0.215	0.262*
8 Months	0.279*	0.271	0.226*
9 Months	0.257*	0.253	0.232*
10 Months	0.299*	0.251*	0.237*
11 Months	0.468*	0.304*	0.323*
12 Months	0.414*	0.358*	0.278*

Source: Authors' estimation

In the short term, within the first five months, moderate correlations are observed between sentiment and tourism arrivals. Significant correlations are only found in the first and fifth months, indicating that the immediate response to media coverage is limited.

In the medium term (6 to 9 months), the correlations grow stronger and become more statistically significant, indicating that media sentiment starts to exert a more noticeable influence on tourism after some delay. The most pronounced effects appear in the long term (10 to 12 months), where correlations reach their highest levels, particularly for overall article sentiment. At 11 months, the correlation coefficient for full article sentiment rises to 0.46, suggesting that sustained positive media coverage has its greatest impact on tourism nearly a year later. This underscores the importance of long-term media narratives in shaping travel decisions.

Furthermore, when comparing different parts of the articles, full article sentiment consistently demonstrates stronger correlations with tourist arrivals than either headlines or opening paragraphs. This suggests that the comprehensive tone and depth of the entire article play a more significant role in influencing tourist perceptions than brief summaries or headlines alone.

Conclusion

It is important that tourism industry service providers comprehend how tourists perceive visiting country. This understanding can influence how potential travelers make their travel decisions that in turn affects tourist numbers and, consequently, the local economy.

This study used sentiment analysis of the UK based international news to understand the changes in Sri Lanka's DI. It uses statistical machine learning models and deep learning models that allowed for a thorough evaluation of sentiment present in international news media.

Among the evaluated models, BERT achieved the best performance, with a weighted F1 score of 0.91, and was therefore chosen for further analysis.

Dataset could be considered as bias, if the results were to be generalized to audiences such as non-English speaking and non-frequent mainstream media. However, findings reveal that sentiment influences tourism arrivals with a time lag, with its impact strengthening over time. The most significant relationships occur between sentiment scores and tourist arrivals 10 to 12 months later, indicating that sustained positive or negative media sentiment can have long-term effects on tourism demand. Additionally, the stronger influence of full article sentiment compared to headlines and opening paragraphs suggests that travelers are more influenced by detailed content than by brief summaries. This emphasizes the importance of comprehensive and positive media coverage in shaping travel behavior.

Although the study offers valuable insights into managing Sri Lanka's destination image (DI), there are still opportunities for further research and refinement. For instance, predictive modeling could be applied to deepen the understanding of the relationship between destination image and tourist arrivals. Moreover, incorporating additional data sources, such as tourist review websites, could enhance the effectiveness and scope of future studies.

Overall, building on the insights from this research can further advance sentiment analysis applications and contribute to the wider field of tourism studies.

References

Ain, Q., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). Sentiment Analysis using Deep Learning Techniques: A Review. *International Journal of Advanced Computer Science and Applications, 8*(6), 424-433. https://doi.org/10.14569/IJACSA.2017.080657

Arachchi, R., & Gnanapala, W. (2020). Impact of COVID 19 and the Future of Tourism Employments in Sri Lanka: A Route to Recovery and Resilience. *Journal of Management and Tourism Research Special Issue*, 15-26.

Bae, Y., & Lee, H. (2012). Sentiment Analysis of Twitter Audiences: Measuring the Positive on Negative Influence of Popular Twitterers. *Journal of the American Society for Information Science and Technology, 63*(12). https://doi.org/10.1002/asi.22768

Chen, H., Huang, X., & Zhiyong, L. (2020). A content analysis of Chinese news coverage on COVID-19 and tourism. *Current Issues in Tourism*, 198-205. doi: https://doi.org/10.1080/13683500.2020.1763269

Chia, S. K.-S., Lo, M.-C., Razak, Z., Wang, Y., & Mohamad, A. A. (2021). Impact of Destination Image on Tourist Satisfaction: The Moderating Effect of Information Technology (IT). *GeoJournal of Tourism and Geosites*, 88-93. https://doi.org/10.30892/gtg.34112-623

Chiorrini, A., Diamantini, C., Mircoli, A., & Potena, D. (2021, March). Emotion and sentiment analysis of tweets using BERT. In *EDBT/ICDT workshops* (Vol. 3, pp. 1-7). [http://star.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-2841/DARLI-AP_17.pdf]

Crompton, J. (1979). Motivations for Pleasure Vacation. *Annals of Tourism Research, 6*(4), 408-424. https://doi.org/10.1016/0160-7383(79)90004-5

De Nisco, A., Mainolfi, G., Marino, V., & Napolitano, M. (2015). Tourism satisfaction effect on general country image, destination image, and post-visit intentions. *Journal of Vacation Marketing, 21*(4), 1-13. https://doi.org/10.1177/1356766715577502

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018, October 11). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.* arXiv.org: https://arxiv.org/abs/1810.04805

Dwyer, L., Gill, A., & Seetaram, N. (2012). *Handbook of Research Methods in Tourism: Quantitative and Qualitative Approaches.* Edward Elgar Publishing.

Fabrice, C., & Pavel, B. (2006). Comparison of SVM and some older classification algorithms in text classification tasks. *Artificial Intelligence in Theory and Practice*, 169-178.

Februadi, A., Wibisono, N., & Purnamasari, D. (2019). Bandung's Image as a Tourist Destination: An Application of Quantitative and Qualitative Approach. *International Journal of Applied Business Research*, 58-69. https://doi.org/10.35313/ijabr.v1i01.41

Gartner, W. C. (1993). Image Formation Process. *Journal of Travel & Tourism Marketing*, 191-216. https://doi.org/10.1300/J073v02n02_12

Ghazali, R., & Cai, L. (2013). Social Media Sites in Destination Image Formation. *Tourism Social Media: Transformations in Identity, Community and Culture*, 73-86. https://doi.org/10.1108/S1571-5043(2013)0000018007

Govers, R., Go, F., & Kumar, K. (2007). Virtual destination image: A new measurement. *Annals of Tourism Research*, 977-997. https://doi.org/10.1016/j.annals.2007.06.001

Guerrero-Rodriguez, R., Alvarez-Carmona, M., Aranda, R., & Lopez-Monroy, A. (2021). Studying Online Travel Reviews related to tourist attractions using NLP methods: the case of Guanajuato, Mexico. *Current Issues in Tourism, 26*(2), 289-304. https://doi.org/10.1080/13683500.2021.2007227

Indra, S., Wikarsa, L., & Turang, R. (2016). Using logistic regression method to classify tweets into the selected topics. *2016 International Conference on Advanced Computer Cience and Information Systems*, 385-390.

Jiang, Q., Chan, C.-S., Eichelberger, S., Ma, H., & Pikkemaat, B. (2021). Sentiment analysis of online destination image of Hong Kong held by mainland Chinese tourists. *Current Issues in Tourism*, 2501-2522. https://doi.org/10.1080/13683500.2021.1874312

Karthika, P., Murugeswari, R., & Manoranjithem, R. (2019). Sentiment Analysis of Social Media Network Using Random Forest Algorithm. *2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, 1-5. https://doi.org/10.1109/INCOS45849.2019.8951367

Kumarasiri, A., Hakeem, R., Karunanayake, R., Senaratne, R., Fonseka, S., Sumanthiran, M., Wickramaratne, J. (2019). *Report of the Select Committee of Parliament to look into and report to Parliament on the Terrorist Attacks that took place in different places in Sri Lanka on 21st April 2019.* Colombo: Department of Government Printing.

Lango M. (2019). Tacking the problem of class imbalance in multi-class sentiment classification: an experimental study. *Foundations of Computing and Decision Sciences, 44*(2), 151-178. https://doi.org/10.2478/fcds-2019-0009

Liu, B. (2022). *Sentiment Analysis and Opinion Mining.* Springer Nature.

Liu, M., Liu, Y., Mo, Z., & Ng, K. L. (2020). Using text mining to track changes in travel destination image: the case of Macau. *Asia Pacific Journal of Marketing and Logistics, 33*(2), 371-393. https://doi.org/10.1108/APJML-08-2019-0477

Mitchell, J. (2022). *Sri Lanka Strategic Plan for Tourism 2022-2025*. Colombo: World Bank.

Nielsen, K. R., Schulz, A., Fletcher, R., & Robertson, T. C. (2025). *The BBC is under scrutiny. Here's what research tells about its role in the UK.* Oxford: Reuters Institute.

SpaCy. (2024). *Industrial-Strength Natural Language Processing*. SpaCy: https://spacy.io/usage

Sri Lanka Tourism Development Authority. (2021). A*nnual Statistical Report 2020. Colombo: Sri Lanka Tourism Development Authority.*

Stepchenkova, S., & Eales, J. (2011). Destination Image as Quantified Media Messages: The Effect of News on Tourism Demand. *Journal of Travel Research*, 198-212. https://doi.org/10.1177/0047287510362780

Tseng, C., Wu, B., Morrison, A. M., Zhang, J., & Chen, Y.-c. (2015). Travel blogs on China as a destination image formation agent: A qualitative analysis using Leximancer. *Tourism Management*, 347-358. https://doi.org/10.1016/j.tourman.2014.07.012

Van Rossum, G. (2020). *The Python Library Reference, release 3.8.2*. Python Software Foundation.

Wang, Z., Udomwong, P., Fu, J., & Onpium, P. (2023). Destination image: A review from 2012 to 2023. *Cogent Social Sciences, 9*(1). https://doi.org/10.1080/23311886.2023.2240569

Zhang, H., Fu, X., Cai, L., & Lu, L. (2014). Destination image and tourist loyalty: A meta-analysis. *Tourism Management*, 213-223. https://doi.org/10.1016/j.tourman.2013.06.006

Zyte. (2023). *Scrapy 2.11 documentation*. Scrapy: https://docs.scrapy.org/en/latest/