International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n1.09
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Exploring Guilin's Destination Image Via User-Generated Content: A Computational Methods

Jianbin Zhu [1, 2], Yuqi Liu [1, *]

[1] College of Foreign Studies, Guilin University of Technology, Guangxi, China
[2] School of Humanities and Education, Foshan University, Guangdong, China

ABSTRACT

This study quantify mainland China's destination image through international tourists' user-generated content (UGC), with a specific focus on Guilin, utilizing advanced computational techniques such as Latent Dirichlet Allocation (LDA) for topic modeling and sentiment analysis. It systematically evaluates online travel reviews to distill key themes and emotional sentiments that shape Guilin's image, uncovering a rich tapestry of cognitive perceptions and affective reactions related to its natural scenery, cultural engagements, service standards, and logistical facets. The application of topic modeling and sentiment analysis provides a nuanced understanding of the cognitive and affective dimensions defining Guilin's destination image. While positive sentiments largely highlight the region's aesthetic and experiential allure, negative sentiments reveal critical areas for improvement, such as perceived value and infrastructure, which are vital for enhancing tourist satisfaction and reinforcing Guilin's appeal in the global tourism market. This research applies LDA and sentiment analysis to interpret UGC offer a calculable methodological approach for broader application in destination image studies. By aligning perceptual and emotional insights with destination marketing strategies, the findings offer actionable intelligence to optimize tourist experiences and enhance destination loyalty. This approach not only enriches the academic understanding of destination images but also provides practical frameworks for destination managers to harness the full potential of UGC in shaping and refining the global image of tourism locales.

KEYWORDS

Guilin; Tourism destination image; User generated contents; Topic modeling; Sentiment analysis

1. INTRODUCTION

Post-COVID-19, Mainland China has rapidly ascended as a key player in the global tourism sector, showcasing emerging trends that cater to both domestic and international markets. The resurgence in domestic tourism, coupled with a burgeoning interest in diverse leisure activities, underscores a market ripe with opportunities (China Briefing, 2023). This period has also witnessed a marked rise in digital engagement and a quest for unique experiences among tourists. The government's commitment to fostering cultural and wellness tourism further exemplifies the dynamic nature of the industry. Concurrently, the concept of "destination image," crucial since its introduction in the 1970s, continues to significantly influence tourist perceptions and choices, underscoring the ongoing challenge for tourism providers to adeptly manage and enhance these perceptions to bolster visitation and loyalty.

The proliferation of online social media and we-media, facilitated by mobile technology, has birthed a vast repository of user-generated content (UGC), including travel logs, star ratings, and tourist

reviews of various destinations. This digital evolution offers unprecedented insights into tourists' experiences and perceptions, transforming tourism research by enabling analyses beyond traditional surveys. Despite its significance, there remains a notable gap in applying advanced computational methods to study mainland China's destination image through international tourists' UGC, even as studies like Hosany (2013) have begun exploring the connections between tourist emotions, satisfaction, and destination evaluations. However, in conducting studies on tourism, researchers often rely on questionnaires and interviews to collect data. While effective for examining specific dynamics and causal relationships, these traditional methods encounter limitations such as sampling bias and temporal restrictions associated with survey administration, challenging the generality and timeliness of findings.

This research aims to fill this gap by leveraging state-of-the-art machine learning techniques, namely Latent Dirichlet Allocation (LDA) for topic modeling and sentiment analysis, to explore the cognitive and affective dimensions of the destination image of Guilin, a renowned tourist destination in China. By analyzing a large datasets of online travel reviews, this study seeks to uncover the unique attributes that contribute to Guilin's image as perceived by international tourists. Specifically, the research focuses on identifying the key themes that emerge from tourists' narratives and the emotional valence associated with their experiences.

The methodology adopted in this study combines quantitative content analysis with advanced natural language processing (NLP) techniques to systematically analyze the content and sentiment of the collected UGC. This approach allows for a nuanced exploration of the multifaceted nature of destination image, encompassing both the cognitive aspects (i.e., knowledge and beliefs about the destination's features) and the affective aspects (i.e., feelings towards the destination). By integrating topic modeling with sentiment analysis, the research aims to provide a comprehensive understanding of the factors that influence Guilin's image among international tourists, thereby offering valuable insights for destination marketers and policymakers.

2. LITERATURE REVIEW

2.1. Tourism Destination Image

The concept of "tourism destination image" was first introduced by Hunt in his doctoral dissertation at Colorado State University, defining it as the perceptions of non-residential places (Hunt, 1971). This idea attracted global academic interest, with Crompton (1979) describing the destination image as a combination of beliefs, ideas, and impressions a person holds about a specific destination. This includes beliefs, perceptions, and feelings that influence an individual's attraction evaluation and choice of destination. Similarly, definitions align in viewing destination image as tourist-based, such as subjective impressions of a place, contrasting with marketer-based images like promotional portrayals (Mackay, 2000; Li & Vogelsong, 2006). These mental images, although potentially accurate or false, real or imagined, play a pivotal role in guiding and shaping behaviors. As tourists face an increasing array of destination choices, destination image serves as a cognitive shortcut in decision-making processes (Kotler & Gertner, 2002). Further, scholars suggest that destination image is also a key indicator of post-visit evaluations, emotional connections with places, and even intentions for revisits and place loyalty (Pan et al., 2017). This expanded view highlights the multifaceted impact of destination image on the tourism experience, emphasizing its significance in tourism research and destination marketing strategies. Destination image is understood through two lenses: the perceptual-cognitive dimension, which encompasses the knowledge and beliefs about a destination's features, and the affective dimension, reflecting one's feelings towards the destination (Berlin & Martin, 2004). This dual perspective is crucial for tourism development and the effectiveness of destination marketing, as it significantly influences both the supply and demand sides

of marketing (Chew & Jahari, 2014), highlighting its pivotal role in shaping tourism experiences and decisions.

2.2. User Generated Contents and Recent Studies on Destination Image

Tourist online reviews are increasingly valued by researchers for their brevity, immediacy, and volume. Advances in technology and internet proliferation have made a wealth of information accessible online. In this landscape, online review data, blogs, and other text-based data form a unique big data category in tourism research – online text data (Binge et al., 2020). Initially, surveys were employed in destination image research, offering precise measurement of specific causal relationships through well-designed studies. However, surveys face challenges in representing selected samples due to their specific time and location constraints. In contrast, user-generated content (UGC) in tourism, such as online reviews or travel blogs, circumvents sampling issues, providing researchers opportunities to gather data on specific topics. Although this data may not represent all tourists, younger, educated individuals with internet access are more likely to post online reviews, offering valuable insights into tourist preferences and behaviors. This shift towards leveraging online text data signifies a broader methodological evolution in tourism research, accommodating the dynamic and diverse nature of tourist experiences and feedback (Bethlehem, 2010).

Over the past few decades, an organic combination of text analysis and destination image technology exists and the application of the combination of the technologies of Sentiment Analysis and topic modeling is also very extensive. Scholars have carried out a lot of research on the measurement of scenic spots based on tourists' online texts and, in the era of big data, the emergence of user generated contents such as tourists' online text information brings increased possibilities to this field. Proposing a machine learning method that incorporates PHFA through the combination of the ideas of management and multi-attribute decision theory, Mirza Alian (2021) utilizes social media analytics to examine destination loyalty in nature-based tourism. Through sentiment analysis and text clustering of TripAdvisor reviews, he identified four key loyalty factors: glaciers, waterfalls, lakes/islands, and hiking/trails. These insights offer valuable guidance for enhancing visitor experiences and loyalty, emphasizing the importance of tailored marketing strategies for tourism providers and organizations (Mirza Alian & Halpenny, 2021). Marine-Roig and Huertas (2020) investigates how the 2017 terrorist attack and Catalan sovereignty process affected Barcelona's online destination image through Airbnb reviews. Analysis using natural language processing on over 150,000 reviews revealed that tourists' perceptions remained largely positive, not significantly deterred by these events. This indicates the resilience of Barcelona's tourism image and the potential for social media to provide nuanced insights into destination loyalty (Marine-Roig & Huertas, 2020). Chu and her colleagues leverage the BERT model for sentiment analysis of travel-related user-generated content on platforms like TripAdvisor, focusing on emotional polarity in comments. Findings reveal tourists predominantly express negative sentiments about food, prices, crowding, and hygiene, influencing negative word-of-mouth. Positive sentiments were less frequent, suggesting specific issues for destination management organizations to address for improved tourist satisfaction and competitive advantage (Chu et al., 2022). Luo presents an augmented mining model combining Latent Dirichlet Allocation (LDA) and Probabilistic Hesitant Fuzzy Algorithm (PHFA) to analyze online reviews for Biophagous destination image. Analyzing 24,222 reviews revealed ten main attributes, with three negatively perceived (crowdedness, tourism cost and accommodation). It offers strategic insights for enhancing tourist experiences and destination management (Luo et al., 2023).

Up to date, however, few articles have used two technologies to explore the destination image of mainland China from international tourists. In addition to that, most of the current studies focus on destination images as a whole or only on cognitive images (Kakawin et al., 2020), the research focusing on the affective image is absent. The LDA model was used to identify and analyze the destination’s cognitive image and affective image based on cognitive-affective theory, exploring tourists’ unique image perception attributes. This study combines the LDA model with the sentiment

analysis method based on lexicon approach, and the satisfaction distribution is further explained by analyzing the emotional polarity through the domain lexicon, mining the focus of positive and negative evaluation, identifying the satisfaction changing trend.

3. METHODOLOGY

3.1. Research Questions

In order to address the above mentioned research gaps in the literature, this study aims to answer the following three research questions:

(1) How do international tourists perceive Guilin's destination image through user-generated content, and what are the cognitive and affective dimensions that define this image?

(2) What are the key themes and sentiments expressed in online reviews about Guilin, and how do they contribute to its overall destination image?

(3) How does the analysis of user-generated content enhance our understanding of destination loyalty and satisfaction trends among international tourists visiting Guilin?

3.2. Content Analysis

Content analysis is a method that quantitatively analyzes the content of non-quantitative literature to find out the core content contained in the literature and the goal of any quantitative content analysis is to produce key category counts and measurements of the quantities of other variables (Neuendorf, 2017). This research integrated sentiment analysis with topic modeling technologies to analyze the content of the collected online travel text data and obtained valuable and meaningful data, then explores the behavioral characteristics and hidden affective attitude of international tourists, and finally explains the research phenomenon.

3.3. Topic Modeling

In natural language processing, topic models are a suite of text-understanding algorithms that unearth the latent patterns from large bodies of unstructured text (Anoop & Sreelakshmi, 2023) . LDA is an unsupervised clustering algorithm that can classify large databases of unstructured text and extract potential topic information (Jing et al., 2023). Additionally, it reduces the semantic dimensions of raw text and is particularly useful for summarizing topics from various datasets (Luo & He, 2021). LDA analysis produces two types of probability distributions: text-topic and topic-word. The topic-word distribution is defined by key words and their probabilities within a topic, where a higher probability indicates a stronger association and contribution to the topic (Ghosh & guha, 2013). Since its inception (Blei, 2003), LDA has become a commonly used topic model which can be used to interpret public sentiment variations and further explore the reasons (Tan et al., 2014). With unsupervised machine learning algorithm to analyze the patterns from the unstructured comments, LDA is commonly used in scenarios dealing with limited observations and automatically cluster similarities where qualitative analysis may fail when dealing with large unstructured text data.

3.4. Sentiment Analysis

Sentiment analysis, informed by frame-semantic theory, extracts affective knowledge from text, utilizing natural language processing (NLP), textual analysis, and computational methods to collect and analyze sentiment data (Alma’s et al., 2017; Hussein, 2018). This approach classifies sentiments in textual documents automatically. Two primary methods employed are lexicon filtering and machine learning, each offers distinct advantages for understanding the nuances of sentiment in large datasets. This analytical process provides invaluable insights into public opinion and consumer

attitudes, supporting a wide range of applications from market research to social media monitoring. The lexicon-based method assigns emotional scores to texts using sentiment dictionaries, aiming to identify overall sentiment through weighted calculations (Pan et al., 2019). Most sentiment dictionaries are manually constructed and analyze text at various syntactic levels (Wankhade et al., 2022). Conversely, machine learning methods train sentiment classifiers with labeled data to predict new sentence sentiments, employing algorithms like maximum entropy, naive Bayes, and SVM (Neethu & Rajasree, 2013). Machine learning is considered more reliable but requires extensive data and training (Gonçalves et al., 2013). However, scholars argue that the machine-learning method is more reliable than the lexicon-filtering approach but the machine-learning approach requires a large body of data and long-term training so that the machine has sufficient examples from which to learn. Therefore, the lexicon method is preferred for large-scale text analysis, especially in analyzing UGC tourism reviews(Liu et al., 2019).

3.5. Data Collection and Preprocessing

The data collection and related preparation work for the study are as follows. The first step is data collection. This study embarked on a comprehensive data collection initiative, focusing on the rich reservoir of online reviews to assess Guilin's tourist attractions. Recognizing TripAdvisor's pivotal role since 2002 in highlighting premier tourism establishments based on global traveler feedback and its broad platform for tourists sharing their opinions (Moro et al., 2019), we harnessed this platform's extensive data. By aggregating over 6,408 reviews from more than 30 renowned Guilin attractions, the study captures a wide array of visitor experiences and opinions, laying a robust groundwork for evaluating destination appeal and visitor satisfaction through unstructured comment analysis. With the help of "Bahay" (octopus) web crawler tool, we collected 6,408 online travel reviews by February 2024, mainly posted between 2012 and 2023. Each review provided the reviewer's name, nationality, number of past reviews, and date of stay, enhancing the dataset's richness and enabling a comprehensive analysis of tourist perceptions and satisfaction. This meticulous data collection process lays a solid foundation for a nuanced exploration of tourist evaluations and destination image, promising valuable insights for both academic research and practical destination management. The second step is data preprocessing and cleaning. The basic process of data preprocessing and cleaning is according to the following steps:

(1) Emoticons, non-English words, punctuation marks, special characters, URL links, and repeated UGC in the text are removed

(2) Remove meaningless stop words with stop-word list from NLTK.

(3) Information from tour guides, other organizations and comments about tour guides service and unrelated attractions was excluded because we wanted to focus on real opinions from tourists that are disseminated online

Finally, from the 6408 reviews, 1758 were discarded and the final dataset comprised 4650 reviews containing a total of 334,170 tokens.

4. FINDINGS AND DISCUSSION

4.1. Detected Topics on UGC

The LDA was applied to extract and label the topics of tourist experience across all collected online reviews from top touristic locations. The key to applying the LDA topic model is to determine the optimal number of topics and the effect of topic extraction in the LDA topic model is directly related to the number of potential topics, and the two most common evaluation methods to determine the number of topics are based on coherence (Stevens et al., 2012) or perplexity. In this study we choose coherence score as a standard to determine the number of topics, which refers to the quantitative

calculation of whether the semantic association of words under a topic generated by an LDA is closer. The LDA identified 8 topics based on the coherence model for this dataset because it had the highest coherence score. Figure 1 shows the coherence score for the number of topics returned by the LDA model. However, as the LDA model operates as an unsupervised learning method, researchers are required to manually assign labels to each topic using the top ten highest-probability words for each topic (Aman et al., 2021). and the labeling of each topic was manually conducted by one researcher and then confirmed by the second researcher. We apply pelvis (https://pypi.org/project/pyLDAvis/; accessed on 7 Mar 2024) to create an interactive figure with the 30 terms when λ = 0.6 as recommended in (Sievert & Shirley, 2014), in which the top 30 most relevant terms were displayed according to frequency once a topic was selected. Figure 2 shows one of the eight topics, Topics 3, 4, 7, and 8 distinct from each other, while there is an overlap between the three remaining topics (Topics 1and 2 as well as topic 5 and 6). Figure 3 and 4 displays the fifteen most relevant terms for each topic and figure 5 shows the top 30 words according to their relative weight with Word Cloud.

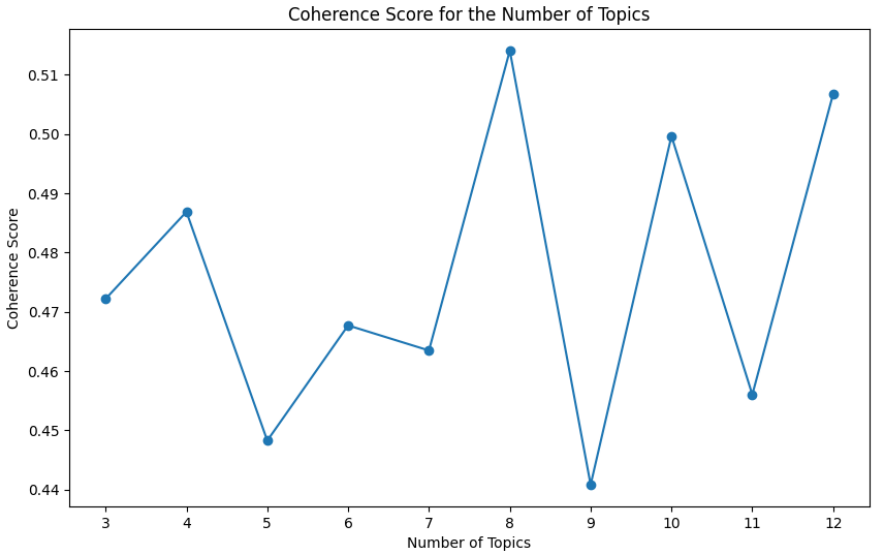

Figure 1. Coherence Scores for Determining Optimal Number of Topics

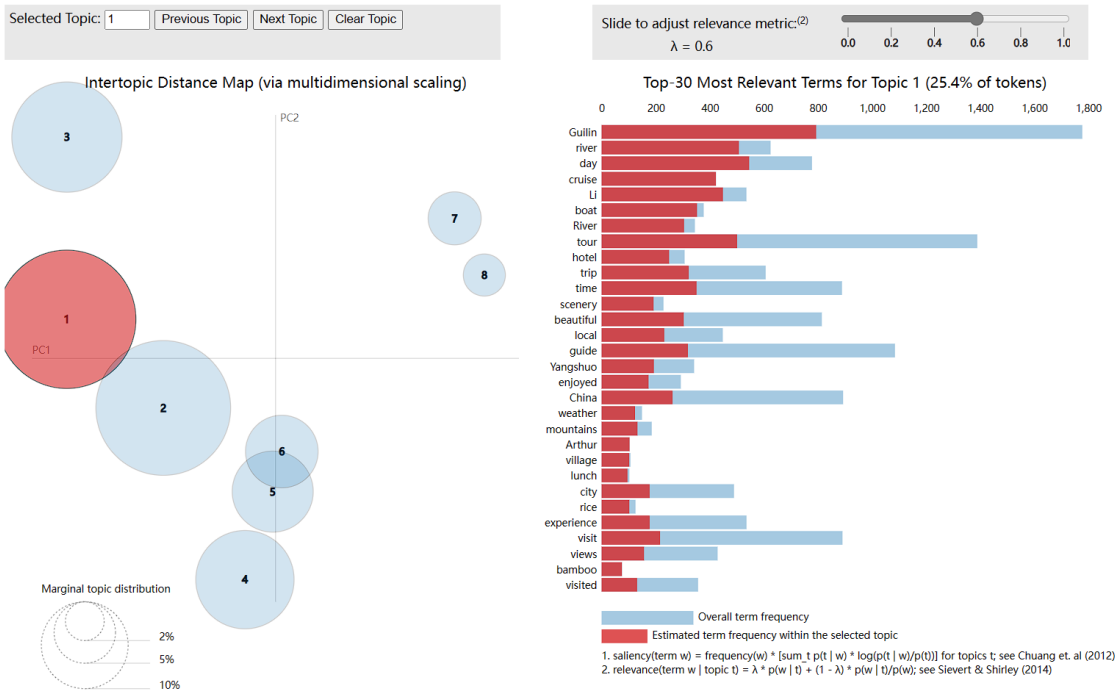

Figure 2. Visual Representation of Topic Distribution with pyLDAvis

Topic 1	Weights 1	Topic 2	Weights 2	Topic 3	Weights 3	Topic 4	Weights 4
ride	0.02312	tea	0.02100	park	0.05008	Guilin	0.02614
mountain	0.02251	zoo	0.01262	hill	0.02258	day	0.01787
car	0.01936	Lu	0.00790	view	0.01844	river	0.01648
top	0.01873	Chinese	0.00671	top	0.01661	tour	0.01614
cable	0.01787	Oscar	0.00576	elephant	0.01492	Li	0.01434
Guilin	0.01689	shop	0.00524	climb	0.01481	cruise	0.01353
view	0.01269	lot	0.00484	nice	0.01458	time	0.01160
views	0.01060	animals	0.00429	walk	0.01384	boat	0.01145
bus	0.01001	pandas	0.00413	Hill	0.01360	trip	0.01056
yuan	0.00782	Tea	0.00409	city	0.01265	guide	0.01021
taxi	0.00737	Japanese	0.00404	views	0.01155	beautiful	0.01018
RMB	0.00711	fun	0.00396	Guilin	0.01101	River	0.00982
walk	0.00706	staff	0.00394	worth	0.01071	China	0.00848
worth	0.00677	Yu	0.00379	entrance	0.00980	hotel	0.00817
Xian	0.00578	Shan	0.00372	Elephant	0.00931	local	0.00761

Figure 3. Top 15 Terms based on their weights for Topics 1-4

Topic 5	Weights 5	Topic 6	Weights 6	Topic 7	Weights 7	Topic 8	Weights 8
tour	0.03139	night	0.03590	cave	0.05442	paintings	0.00608
guide	0.02849	lake	0.02724	caves	0.02266	artist	0.00456
China	0.02688	walk	0.02385	visit	0.01595	package	0.00432
Guilin	0.02189	pagodas	0.02115	lights	0.01440	Dont	0.00387
recommend	0.01164	nice	0.01406	lighting	0.01387	bought	0.00364
trip	0.01098	street	0.01381	formations	0.01216	black	0.00344
Travel	0.01024	lakes	0.01123	worth	0.01009	Journey	0.00330
Guide	0.01002	lit	0.01099	natural	0.00961	tropical	0.00315
experience	0.01000	shops	0.01090	tour	0.00937	mistake	0.00302
time	0.00933	evening	0.01084	beautiful	0.00933	paint	0.00298
days	0.00800	park	0.01080	inside	0.00831	shop	0.00291
knowledgeable	0.00759	town	0.01025	Chinese	0.00822	painting	0.00285
Yangshuo	0.00748	beautiful	0.01002	nice	0.00715	pushy	0.00282
Shanghai	0.00720	restaurants	0.00976	guide	0.00709	enormous	0.00271
guides	0.00705	food	0.00932	Guilin	0.00650	rushed	0.00266

Figure 4. Top 15 Terms based on their weights for Topics 5-8

Figure 5. Word Cloud of Top 30 Terms for Each Topics

4.2. Sentiment Polarity Statistic Based on Domain Lexicon

In this study, we use the NRC lexicon to conduct the sentiment analysis to get the exclusive sentiment lexicon for Guilin tourists reviews, as it has been implemented successfully in similar tourism review research (Feizollah et al., 2021). In addition to classifying data into positive and negative sentiments, the NRC lexicon, which was assembled manually via crowdsourcing, is capable of classifying emotions into eight basic categories: trust, surprise, fear, sadness, anger, joy, anticipation, and disgust. It has been proved that human sentiments could be categorized into specific primary emotions (Wu et al., 2014); therefore, we classified the data on TripAdvisor online reviews into the eight emotion classes from the NRC lexicon.

The NRC lexicon enabled us to classify the tone or emotion of the text into eight categories to gain more insight into the emotion of the users, as opposed to just positive or negative sentiment. The online reviews were classified into fear, disgust, anger, and sadness—the negative emotions—and trust, joy, surprise, and anticipation—the positive emotions. Figure 6 shows the percentage of the corpus that is classified as each emotion. Trust, anticipation and joy were the top three emotions found in the reviews, which are considered positive emotions towards Guilin attractions. Figure 7 and figure 8 shows the top 20 most repeated positive and negative words in UGC. The words “Guilin” and “China” were removed due to their excessive repetition.

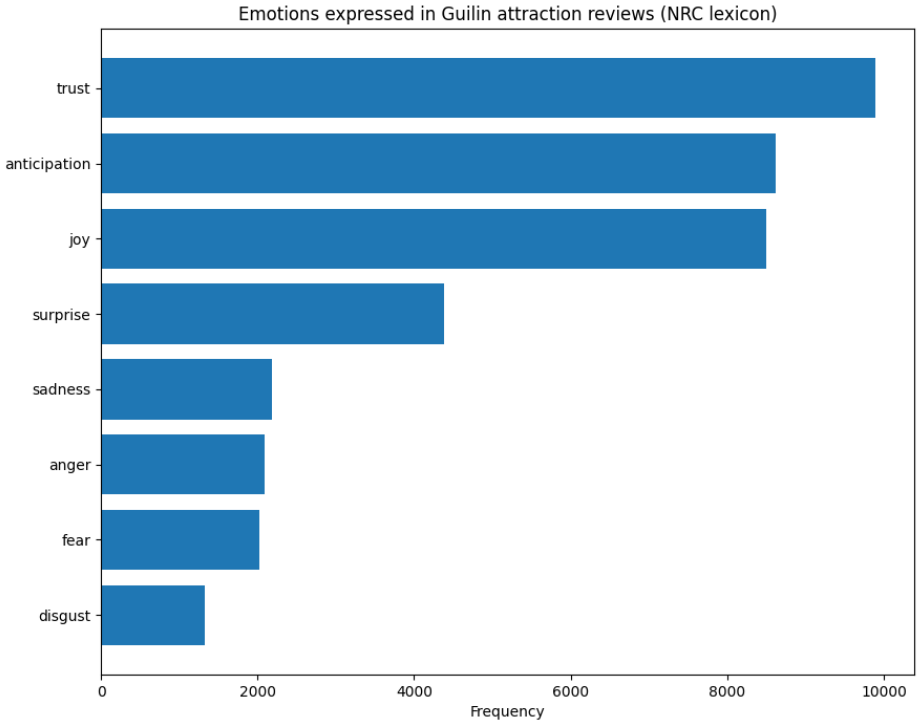

Figure 6. Emotional Distribution of Guilin Tourist Reviews

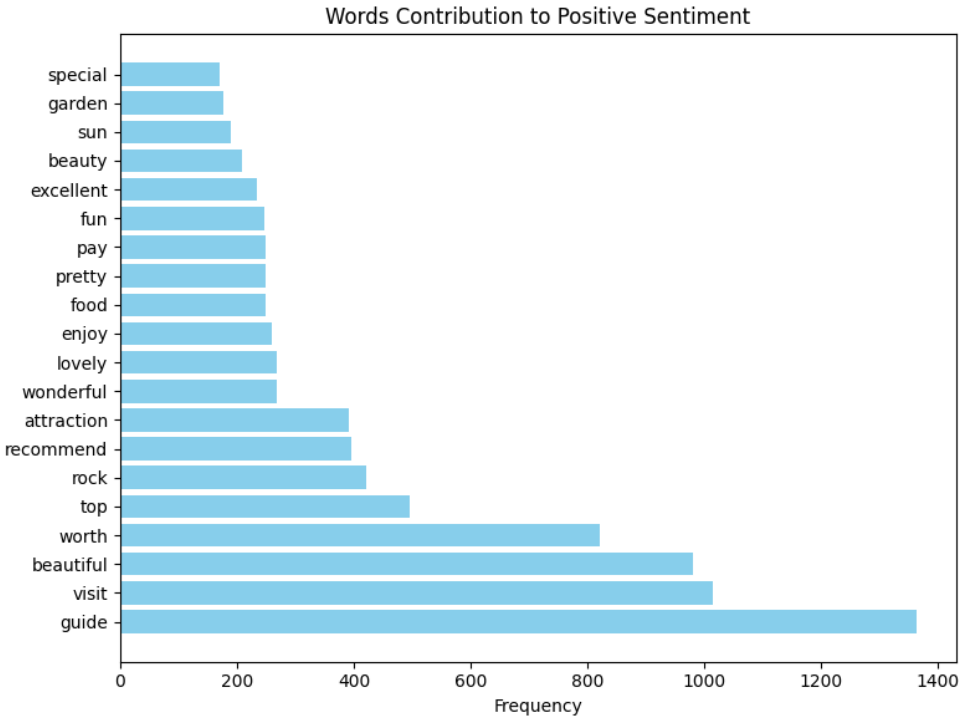

Figure 7. Top 20 Most Repeated Positive Words in User-Generated Content

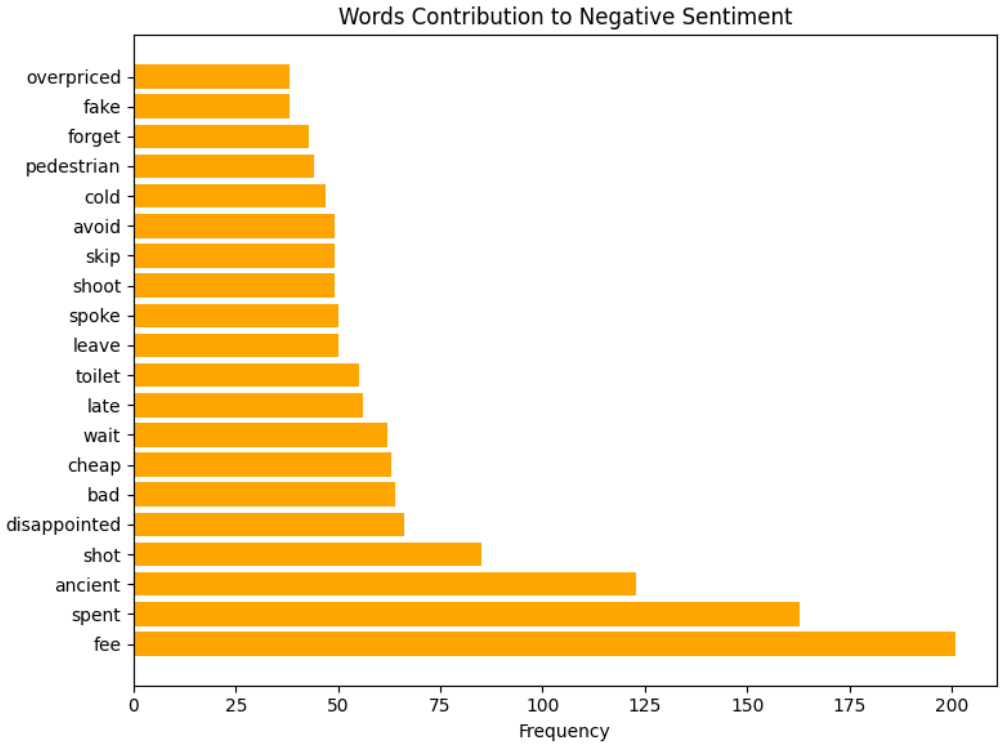

Figure 8. Top 20 Most Repeated Negative Words in User-Generated Content

4.3. Results and Discussion

The Latent Dirichlet Allocation (LDA) analysis of online reviews from international tourists visiting Guilin reveals a multifaceted destination image characterized by a diverse array of attractions and experiences. According to the results of figure 3-5, we manually labeled these eight themes which include picturesque natural scenery and mountains, a blend of zoological and traditional tea experiences, urban parks, river and cruise activities, organized travel with guided tours, vibrant cultural nightlife, explorations of famous caves like the Reed Flute Cave, and engaging artistic and shopping experiences. Each theme underscores aspects of Guilin that contribute to its attractiveness to international tourists, showcasing a wide range of experiences from natural beauty to cultural richness. Tourists are captivated by Guilin's picturesque mountainous terrains and natural beauty, frequently accessed via convenient transportation options like cable cars and buses, which are positively noted for enhancing the scenic exploration of the region. Simultaneously, the unique blend of wildlife encounters and the deep-rooted tea culture draws appreciation, reflecting a harmonious integration of natural and cultural tourism. Urban attractions, particularly parks and sites like Elephant Hill Park, blend urban convenience with natural splendor, although challenges such as steep treks and admission fees occasionally temper the enthusiasm. Water-based activities, including tranquil river cruises, receive high praise for their relaxing and visually appealing experiences, underscoring Guilin’s charm as a water-centric retreat. The structured support of guided tours enriches tourist experiences, with skilled guides providing valuable insights into the local heritage. Guilin’s vibrant nightlife and cultural activities resonate well, with evening strolls around luminous lakes and bustling streets highlighting the city’s dynamic cultural scene. Cave explorations like those in Reed Flute Cave fascinate with stunning geological formations, though confined spaces and artificial lighting might not appeal to everyone. Lastly, while the art and shopping experiences offer interaction with local craftsmanship, they are occasionally marred by aggressive sales tactics, slightly marring an otherwise enriching cultural exchange. Overall, these themes collectively paint Guilin as a destination that boasts a rich tapestry of natural landscapes, cultural vitality, and engaging activities, with minor areas for improvement in visitor experience and commercial practices.

Apart from that, we conducted a keyword concurrence analysis through VOS viewer software based on each term’s association strength in figure 9. VOS-viewer software was developed by the Science and Technology Research Center of Leiden University in the Netherlands (Van & Waltman, 2017). It can provide three visualization methods: network view, superposition view and density view, with the characteristics of beautiful image and easy to interpret. The keyword co-occurrence analysis of reviews from international tourists visiting Guilin reveals four thematic clusters, each distinguished by distinct experiences and attractions. The first green cluster, dedicated to caves, captures the enchantment of Guilin's subterranean landscapes, with visitors particularly drawn to the spectacular formations and interplay of colors and light. However, mentions of entrance fees and brief tours suggest some concerns over value and tour duration. The blue cluster focuses on the serene settings of parks and lakes, where terms like "Elephant Hill," "pagoda," and "scenic area" underscore the seamless integration of natural splendor with human-crafted landscapes, cherished for tranquil strolls and picturesque views, especially noted during night-time reflections. IN the red cluster, discussions revolve around rich cultural and educational experiences, where engagement with local customs, language lessons, and photography are prominent, often enhanced by interactions with notable guides such as "Mia" and "Becky." This cluster highlights the depth of cultural immersion available to tourists. The last yellow cluster illustrates the critical role of guided tours, with praise for the guides' professionalism and knowledge, pointing to a well-appreciated aspect of the tourism experience that deepens visitors' understanding and enjoyment of Guilin.

Collectively, these clusters not only portray Guilin as a multifaceted destination rich in both natural beauty and cultural depth but also pinpoint areas for enhancement. Particularly, the management of expectations regarding tour logistics and cost could be improved. Addressing these concerns could help in refining Guilin's image as a world-class tourist destination, ensuring that it not only meets but exceeds the expectations of its international visitors, thus fostering a more satisfying and holistic travel experience.

Lastly, according to the sentiment analysis of Guilin attraction reviews using the NRC lexicon, it provides a comprehensive look into the emotional responses of international tourists, revealing an overwhelmingly positive emotional landscape dominated by sentiments of 'trust,' 'anticipation,' and 'joy.' This indicates that tourists generally leave with high levels of satisfaction, enjoying both the scenic beauty and the enriching experiences Guilin has to offer. The prominence of joy is significantly marked by frequent references to the region's picturesque landscapes and the exceptional quality of tours, where terms such as "beautiful," "visit," and "guide" often surface, pointing towards the outstanding natural beauty of Guilin and the effectiveness of tour guides in enhancing visitor experiences. Moreover, the use of words like "wonderful," "lovely," and "recommend" in the reviews not only contributes to a strong positive sentiment but also suggests that visitors are likely to endorse Guilin as a destination to others, underlining the area's appeal and the likelihood of repeat visits. This level of endorsement is crucial for sustaining and growing the tourist influx, which is vital for the local economy.

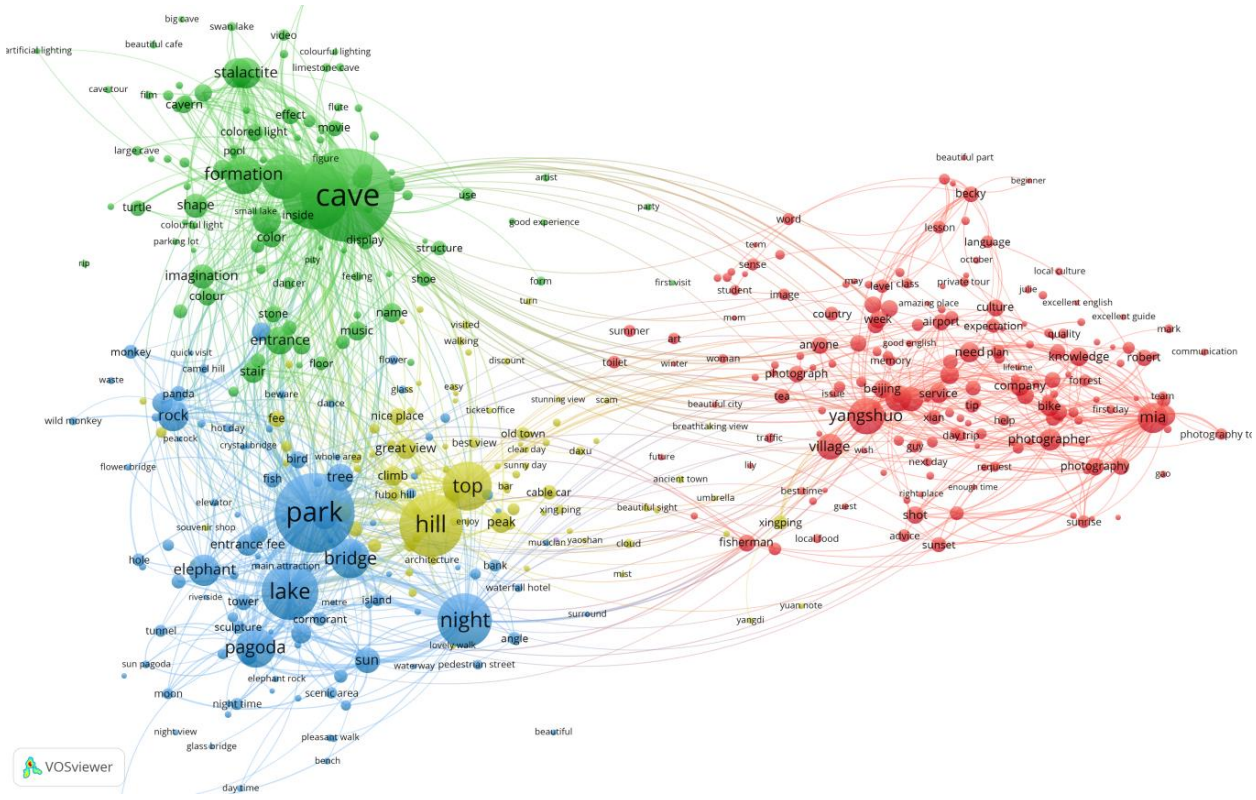

Figure 9. Keyword Co-Occurrence Analysis Visualization

However, the analysis also uncovers a spectrum of negative emotions such as 'sadness,' 'anger,' and 'fear,' which, although less frequent, highlight areas of concern. These negative sentiments often relate to specific disappointments or unmet expectations, which, if addressed, could significantly enhance the tourist experience. Specific words like "overpriced," "disappointed," and "fee" suggest that some tourists perceive certain attractions or services as not providing sufficient value for money. Additionally, logistical issues reflected in words like "wait" and "toilet" indicate dissatisfaction with some practical aspects of the tourism infrastructure in Guilin.

Addressing these negative sentiments by focusing on areas such as pricing transparency, enhancing facility quality, and improving overall visitor management could mitigate the concerns expressed by tourists. By refining these elements, Guilin could further solidify its reputation as a premier tourist destination, ensuring that the experiences provided not only meet but exceed the expectations of international visitors, fostering a more satisfying and holistic travel experience. This strategic focus on continuous improvement will help maintain Gauguin's attractiveness as a top-notched destination worldwide, encouraging longer stays and more enthusiastic recommendations.

5. CONCLUSION

Through a detailed examination of thematic models, keyword co-occurrences, and sentiment analyses derived from tourist reviews, this study provides a nuanced understanding of how international tourists perceive Guilin's destination image. The research reveals both the cognitive and affective dimensions that define this image, demonstrating how key themes and sentiments expressed in online reviews contribute to shaping Guilin's overall destination image.

On the one hand, the perception of Guilin through user-generated content reveals a complex blend of cognitive and affective dimensions that shape its image. Tourists appreciate Guilin's natural and cultural attributes, evidenced by their emotional responses dominated by trust, joy, and anticipation. These emotions reflect a deep appreciation for both the aesthetic beauty and the enriching experiences Guilin provides, pointing to a strong affective connection with the destination. The key themes

emerging from the reviews—ranging from natural scenery and cultural encounters to service quality and logistical aspects—highlight the factors that contribute to Guilin's image. Positive sentiments predominantly underline the beauty and unique experiences Guilin offers, while negative sentiments, though fewer, point to areas needing improvement such as value for money and infrastructure. These themes are crucial as they directly influence tourists' overall perception and the narrative they share with potential visitors and the global tourism market.

However, despite its contributions, this study is not without limitations and deficits. First, the reliance on user-generated content, while insightful, may not fully represent the broader tourist population due to selection bias, as it primarily reflects the views of those motivated to post reviews online. In addition to, the analysis is based on textual data from reviews, which can be challenging to interpret accurately across different languages and cultural contexts, potentially leading to inaccuracies in sentiment analysis. Lastly, the focus on publicly available online reviews excludes data from private feedback channels or lesser-known review platforms, which could provide additional insights into tourist experiences. These limitations highlight the need for future research to incorporate more diverse data sources and develop advanced methods to address linguistic and cultural variations in user-generated content.

In conclusion, this multifaceted analysis of user-generated content not only answers how international tourists perceive and experience Guilin but also elucidates the broader implications of these perceptions for destination management and marketing strategies. By leveraging these insights, stakeholders can enhance Guilin's appeal and competitiveness in the international tourism market, ensuring it remains a favored choice among global travelers.

References

[1] Anoop, V. S., & Sreelakshmi, S. (2023). Public discourse and sentiment during Mpox outbreak: An analysis using natural language processing. Public Health, 218, 114–120. https://doi.org/10.1016/j.puhe.2023.02.018

[2] Aman, J. J. C., Smith-Colin, J., & Zhang, W. (2021). Listen to E-Scooter Riders: Mining Rider Satisfaction Factors from App Store Reviews. Transportation Research Part D: Transport and Environment, 95(June), 102856. https://doi.org/10.1016/j.trd.2021.102856

[3] Almars, A., Li, X., Zhao, X., Ibrahim, I.A., Yuan, W.W. and Li, B.H. (2017), "Structured sentiment analysis", in Cong, G., Peng, W.C., Zhang, W.E., Li, C. and Sun, A. (Eds), Advanced Data Mining and Applications, Adma 2017, pp. 695-707.

[4] Beerli, A. and Martin, J.D. (2004), "Factors influencing destination image", Annals of Tourism Research, Vol. 31, pp. 657-681.

[5] Bethlehem, J. (2010). Selection bias in web surveys. International Statistical Review, 78, 161 – 188.

[6] Bigne, E., Zanfardini, M. and Andreu, L. (2020), "How online reviews of destination responsibility influence tourists' evaluations: an exploratory study of mountain tourism", Journal of Sustainable Tourism, Vol. 28, pp. 686-704.

[7] Blei, D. M. (unknown). Latent Dirichlet Allocation.

[8] Chu, M., Chen, Y., Yang, L., & Wang, J. (2022). Language interpretation in travel guidance platform: Text mining and sentiment analysis of TripAdvisor reviews. Frontiers in Psychology, 13, 1029945. https://doi.org/10.3389/fpsyg.2022.1029945

[9] China Briefing. (2023). China Tourism in 2023: Outlook, Trends and Opportunities. Retrieved from www.china-briefing.com

[10] Chew, E. and Jahari, S. A. (2014), "Destination image as a mediator between perceived risks and revisit intention: a case of post-disaster Japan", Tourism Management, Vol. 40, pp. 382-393.

[11] Feizollah, A., Mostafa, M. M., Sulaiman, A., Zakaria, Z., & Firdaus, A. (2021). Exploring halal tourism tweets on social media. Journal of Big Data, 8(1), 72. https://doi.org/10.1186/s40537-021-00463-5

[12] Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2013). Comparing and Combining Sentiment Analysis Methods. Proceedings of the First ACM Conference on Online Social Networks, 27–38. https://doi.org/10.1145/2512938.2512951

[13] Ghosh, D.D.; Guha, R. What are we 'tweeting' about obesity? Mapping tweets with Topic Modeling and Geographic Information System. Cartogr. Geogr. Inf. Sci. 2013, 40, 90–102. [CrossRef]

[14] Hussein, D.M.E.-D.M. A survey on sentiment analysis challenges. J. King Saud Univ. Eng. Sci. 2018, 30, 330–338.

[15] Hunt J D. Image As a Factor in Tourism Development [D]. Colorado State University, 1971.

[16] Hosany, S., & Gilbert, D. (2010). Measuring tourists' emotional experiences toward hedonic holiday destinations. Journal of Travel Research, 49(4), 513e526.

[17] Jing, P., Cai, Y., Wang, B., Wang, B., Huang, J., Jiang, C., & Yang, C. (2023). Listen to social media users: Mining Chinese public perception of automated vehicles after crashes. Transportation Research Part F: Traffic Psychology and Behaviour, 93, 248–265. https://doi.org/10.1016/j.trf.2023.01.018

[18] Kiatkawsin, K., Sutherland, I., & Kim, J.-Y. (2020). A Comparative Automated Text Analysis of Airbnb Reviews in Hong Kong and Singapore Using Latent Dirichlet Allocation. Sustainability, 12(16), 6673. https://doi.org/10.3390/su12166673

[19] Kotler, P., Gertner, D. Country as brand, product and beyond: A place marketing and brand management perspective [J]. Journal of Brand Management, 2002, 9(4–5):249~261.

[20] Luo, S., & He, Sylvia Y. (2021). Understanding Gender Difference in Perceptions toward Transit Services across Space and Time: A Social Media Mining Approach. Transport Policy, 111(September), 63–73. doi:10.1016/j.tranpol.2021.07.018.

[21] Li X, Vogelsong H. Comparing methods of measuring image change: A case study of a small-scale community festival [J]. Tourism Analysis, 2006, 10(4): 349~360.

[22] Liu, Y., Huang, K., Bao, J., & Chen, K. (2019). Listen to the voices from home: An analysis of Chinese tourists' sentiments regarding Australian destinations. Tourism Management, 71, 337–347. https://doi.org/10.1016/j.tourman.2018.10.004

[23] Luo, Y., Tong, T., Zhang, X., Yang, Z., & Li, L. (2023). Exploring destination image through online reviews: An augmented mining model using latent Dirichlet allocation combined with probabilistic hesitant fuzzy algorithm. Kybernetes, 52(3), 874–897. https://doi.org/10.1108/K-07-2021-0584

[24] Marine-Roig, E., & Huertas, A. (2020). How safety affects destination image projected through online travel reviews. Journal of Destination Marketing & Management, 18, 100469. https://doi.org/10.1016/j.jdmm.2020.100469

[25] Moro, S.; Esmerado, J.; Ramos, P.; Alturas, B. Evaluating a guest satisfaction model through data mining. Int. J. Contemp. Hosp. Manag. 2019, 32, 1523–1538. [CrossRef]

[26] Mackay, K.J., Fesenmaier, D.R. An exploration of cross-cultural destination image assessment [J]. Journal of Travel Research, 2000, 83(4):417~423.

[27] Mirzaalian, F., & Halpenny, E. (2021). Exploring destination loyalty: Application of social media analytics in a nature-based tourism setting. Journal of Destination Marketing & Management, 20, 100598. https://doi.org/10.1016/j.jdmm.2021.100598

[28] Neethu, M. S., & Rajasree, R. (2013). Sentiment analysis in twitter using machine learning techniques. 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 1–5. https://doi.org/10.1109/ICCCNT.2013.6726818

[29] Neuendorf, K. A. (2017). The content analysis guidebook (2n ed.). London: SAGE Publications.

[30] Pan, D., Yuan, J., Li, L., & Sheng, D. (2019). Deep neural network-based classification model for Sentiment Analysis. 2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC), 1–4. https://doi.org/10.1109/BESC48373.2019.8963171

[31] Pan, L.; Zhang, M.; Gursoy, D.; Lu, L. Development and validation of a destination personality scale for mainland Chinese travelers. Tour. Manag. 2017, 59, 338–348.

[32] Stevens, K.; Kegelmeyer, P.; Andrzejewski, D.; Buttler, D. Exploring topic coherence over many models and many topics. In Proceedings of the Empirical Methods in Natural Language Processing 2012, Jeju Island, Korea, 12–14 July 2012.

[33] Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, 63–70. https://doi.org/10.3115/v1/W14-3110

[34] Tan, S., Li, Y., Sun, H., Guan, Z., Yan, X., Jiajun, B.u., et al. (2014). Interpreting the Public

[35] Van Eck, N.J., Waltman, L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics, 2017, 111(2): 1053-1070.

[36] Wu Y, Yang Y, Chiu C-Y. Responses to religious norm defection: the case of Hui Chinese Muslims not following the halal diet. Int J Intercult Relat. 2014; 39:1–8.

[37] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1