Journal of Information Systems and Informatics
Vol. 6, No. 2, June 2024 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v6i2.743 Published By DRPM-UBD

# Understanding Visitor Sentiment of Batu Cave Destination through TripAdvisor and Vlogger Content Reviews

**Yerik Afrianto Singgalen[1*]**

[1*] Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: [1*]yerik.afrianto@atmajaya.ac.id

**Abstract**

This study utilizes the CRISP-DM framework to conduct a comprehensive sentiment analysis of visitor reviews for Batu Cave, leveraging advanced tools such as VADER, TextBlob, and the SVM model. The analysis of 1201 TripAdvisor reviews reveal critical visitor perceptions, highlighting both positive aspects, such as the site's beauty and cultural significance, and areas needing improvement, including accessibility and visitor conduct. The SVM model demonstrates high performance with an accuracy of 94.25% and AUC scores of 0.966 (optimistic), 0.962 (standard), and 0.958 (pessimistic). Furthermore, toxicity scores from the Perspective API range from 0.05055 to 0.89882, identifying areas for enhancing visitor interactions. These findings underscore the importance of using data-driven approaches to improve destination management and visitor satisfaction. The study provides valuable insights for policymakers, guiding strategic planning and sustainable development of tourist destinations. Consequently, the research offers a robust foundation for informed decision-making in the tourism sector, aiming to enhance the overall visitor experience at Batu Cave.

**Keywords**: Batu Cave; CRISP-DM; Malaysia; Sentiment; TripAdvisor

## 1. INTRODUCTION

Visitor sentiment analysis constitutes a critical study area to ascertain tourists' perceptions, motivations, and preferences when visiting a destination. This analysis provides valuable insights into visitors' emotional responses and satisfaction levels, enabling destination managers to tailor services and marketing strategies [1]–[3]. Understanding these factors is essential for enhancing the visitor experience, as it allows for the identification of critical areas requiring improvement and innovation [4]–[8]. Consequently, such an approach fosters a more appealing and competitive destination, leading to sustained tourist engagement and economic benefits [9]–[12]. Therefore, conducting comprehensive visitor sentiment analysis is indispensable for informed decision-making and strategic planning in the tourism sector.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

In tourism, the study of visitor sentiment is intricately linked to consumer behavior, capturing purchasing patterns and perceptions of products and services during travel. This analysis encompasses understanding how tourists make decisions, what influences choices, and how they evaluate experiences [13]–[17]. Such insights are pivotal for optimizing service quality and product offerings, directly impacting customer satisfaction and loyalty [18]–[22]. Therefore, examining visitor sentiment provides essential data that drives strategic improvements and enhances overall destination appeal, ultimately fostering a sustainable and competitive tourism environment.

Understanding visitor sentiment provides crucial insights for policymakers to develop destination programs that align with tourist preferences, ensuring sustainability. By analyzing sentiments, decision-makers identify critical areas of satisfaction and areas needing improvement, tailoring strategies to meet visitors' expectations more effectively [23]. This alignment enhances the visitor experience and fosters long-term engagement and positive word-of-mouth, which is essential for sustainable tourism growth [24]–[28]. Consequently, integrating sentiment analysis into policy formulation enables a more responsive and adaptive approach to destination management, promoting enduring success and visitor satisfaction.

This study aims to identify and analyze the sentiments of visitors to Batu Caves by processing text data from reviews on TripAdvisor and content from vloggers on YouTube. The investigation systematically examines user-generated content to uncover prevalent emotional responses and overall satisfaction levels associated with the site [29]. Advanced text analysis techniques, such as sentiment analysis and natural language processing, provide a comprehensive understanding of public perception [30]. By leveraging these insights, the study provides valuable implications for tourism management and marketing strategies at Batu Cave. The findings are anticipated to enhance visitor experiences and inform targeted service quality improvements.

The framework employed in the sentiment analysis of Batu Cave's destination is CRISP-DM, a robust methodology for data mining projects. This structured approach facilitates a comprehensive process, beginning with business understanding and progressing through data understanding, modeling, evaluation, and deployment [31]. By following these phases, the analysis ensures a meticulous and replicable examination of visitor sentiments derived from diverse data sources [32]. The systematic application of CRISP-DM is believed to enhance the accuracy and reliability of the sentiment analysis, providing actionable insights for improving the management and promotion of Batu Cave. Consequently, this framework offers a valuable blueprint for future sentiment analysis endeavors in the tourism sector.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

The research on sentiment analysis of Batu Cave visitors holds significant theoretical and practical implications. Theoretically, it advances the understanding of sentiment analysis methodologies within tourism, contributing to the academic discourse on data-driven approaches to gauge visitor experiences [33]–[37]. Practically, the insights derived from this study offer valuable guidance for tourism stakeholders to enhance service quality and marketing strategies [38]–[41]. This dual impact underscores the importance of integrating sentiment analysis into tourism studies, as it bridges the gap between academic exploration and real-world application, ultimately leading to more informed decision-making processes in the industry.

Similar research on sentiment analysis in tourism has yielded valuable insights but also faces notable limitations. Studies examining visitor sentiments at various tourist destinations have successfully utilized text-mining techniques to capture public perceptions and enhance service offerings [42]. However, challenges such as data heterogeneity, language nuances, and contextual variability often hinder the accuracy and generalizability of findings. Addressing these limitations requires methodological advancements and the inclusion of diverse data sources to ensure comprehensive and reliable outcomes. Therefore, while sentiment analysis remains a powerful tool in tourism research, its application must continuously evolve to overcome inherent constraints and deliver actionable insights.

## 2. METHODS

### 2.1 Gap Analysis

A literature review was conducted to identify gaps in studies related to tourist sentiment and tourist behavior. This comprehensive analysis involved examining existing research to uncover areas that have been underexplored or insufficiently addressed. By highlighting these gaps, the review provides a foundation for future investigations to build upon and expand the current understanding of tourist dynamics. Addressing these identified gaps is crucial for developing more nuanced and effective tourism management and marketing strategies. Ultimately, this literature review is pivotal in advancing the field by directing attention to overlooked aspects and promoting more holistic research approaches.

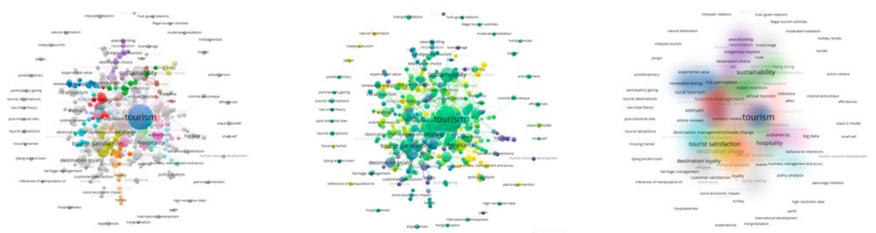


**Figure 1.** Network, Overlay, and Density Visualization of Tourist Behavior

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Figure 1 shows the gap analysis of AI research. A correlation between tourist satisfaction and destination loyalty has been established based on identifying similar research topics. Studies indicate that tourists who express high satisfaction levels after visiting a destination are more likely to exhibit intentions to revisit [43]–[45]. This relationship underscores the importance of enhancing tourist satisfaction to foster destination loyalty [46]–[49]. Enhancing satisfaction improves the immediate visitor experience and contributes to sustained tourism growth through repeat visits and positive word-of-mouth. Therefore, prioritizing tourist satisfaction is critical for achieving long-term destination success.

Identifying research gaps reveals that studies on tourist behavior are effectively conducted using sentiment analysis through the CRISP-DM framework. This approach facilitates a structured examination of tourist sentiments by systematically processing and analyzing text data from various sources. Employing CRISP-DM ensures a comprehensive and replicable methodology, enhancing the reliability of the findings. Integrating sentiment analysis with this framework offers a robust tool for uncovering nuanced insights into tourist behavior. Consequently, this method holds significant potential for advancing the understanding of tourist preferences and improving destination management strategies.

### 2.2 Cross-Industry Standard Process for Data-Mining (CRISP-DM)

The CRISP-DM approach emphasizes the contextuality of datasets, ensuring that sentiment analysis is conducted in alignment with the specific context of the data [50]. This study focuses on reviews from TripAdvisor and YouTube content about Batu Cave in Kuala Lumpur, Malaysia, necessitating a tailored analysis to capture the unique sentiments expressed by visitors. Considering the context in which the data was generated, the analysis becomes more accurate and meaningful, reflecting genuine visitor experiences. The application of CRISPR-DM in this context underscores its effectiveness in producing reliable insights that inform targeted improvements in tourism management. Consequently, this approach not only enhances the quality of the analysis but also supports the development of more nuanced and effective strategies for destination enhancement.

Figure 2 shows the implementation of CRISP-DM. Based on the data context, it is evident that the first video is a vlog with the ID 40aSCy-wgM4, while the second video is a vlog with the ID ossQvd4K9VU. These identifiers are crucial for locating and analyzing the specific content relevant to the study. Understanding the distinct sources of these vlogs allows for more targeted sentiment analysis, capturing the nuances of visitor experiences as depicted in these videos. Consequently, accurately referencing these vlog IDs enhances the reliability of the analysis, ensuring that the findings are directly applicable to the observed data. This precision ultimately contributes to a more detailed and contextually relevant understanding of visitor sentiments.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

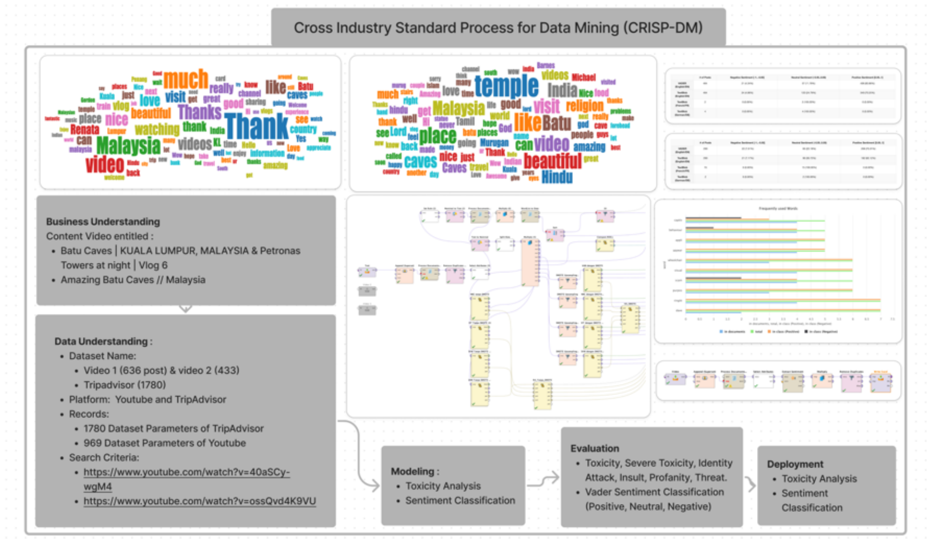



**Figure 2.** Implementation of Cross-Industry Standard Process for Data-Mining (CRISP-DM)

The CRISP-DM framework establishes that the business understanding phase is focused on visitor sentiment at Batu Caves. This initial phase involves defining the objectives and requirements from a business perspective, ensuring that the analysis aligns with the intention to understand visitor experiences comprehensively. By concentrating on visitor sentiment, the framework sets the stage for a targeted and relevant analysis that addresses specific business needs. This approach enhances the clarity and direction of the subsequent analysis phases and ensures that the outcomes directly apply to improving visitor satisfaction and destination management. Consequently, focusing on visitor sentiment in the business understanding phase is pivotal for achieving meaningful and actionable insights.

### 2.2.1 Business Understanding

Based on visitor reviews of Batu Caves on TripAdvisor, it is possible to discern the characteristics of visitors by country of origin, rating, visit time, year of visit, and visitor type. This data provides a comprehensive profile of tourists' diverse demographic and temporal patterns, highlighting satisfaction levels and visit frequency variations. Such detailed segmentation allows for targeted analysis of visitor experiences, revealing trends and preferences unique to different groups. Understanding these characteristics is essential for tailoring tourism strategies to enhance visitor satisfaction and optimize destination management. Consequently, these insights contribute to a more informed approach to improving the overall tourist experience at Batu Caves.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Based on 1201 review data from TripAdvisor, it is evident that reviewers come from a diverse array of countries. The analysis reveals that visitors from London, Singapore, and Melbourne, Australia, constitute most of the reviewers. This geographical diversity highlights the global appeal of Batu Caves as a tourist destination. The prominence of reviews from these specific locations suggests that targeted marketing efforts in these areas could further enhance visitor engagement and satisfaction. Consequently, understanding the origins of the most frequent reviewers provides valuable insights for optimizing tourism strategies and improving the overall visitor experience.

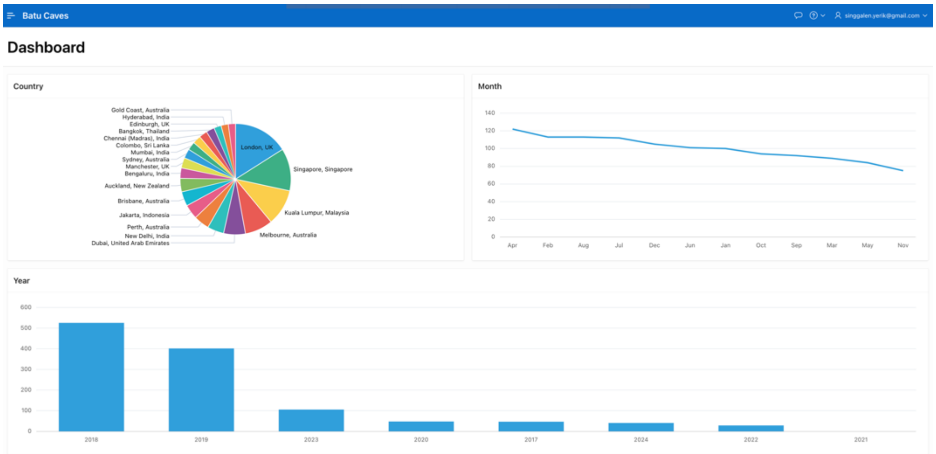


**Figure 3.** Visitor Country, Month, and Year of Visit

Figure 3 shows the visitor's country, month, and year of visit. When classified by month, the highest number of visits occurred in April 2018 and 2019. This peak in visitation suggests a seasonal trend that specific events or favorable weather conditions may influence during this period. The consistent increase in visitor numbers in April over these two years indicates a pattern tourism managers could leverage for planning and promotional activities. Consequently, focusing on the factors contributing to this surge helps enhance tourist experiences and strategically manage visitor flows throughout the year.

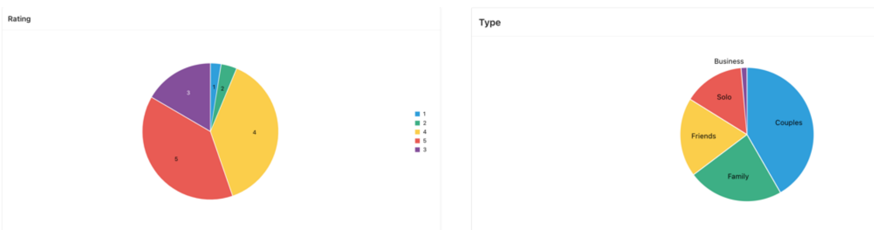



**Figure 4.** Rating and Type of Visitor

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Figure 4 shows the rating and visitor type. Based on visitor ratings for Batu Caves, most reviewers were awarded ratings of 5 and 4, indicating "excellent" and "excellent" experiences, respectively. Following these, most visitors rated the experience as 3, which denotes an "average" experience. This rating distribution highlights a generally positive reception, with most visitors expressing high satisfaction. The prevalence of top ratings suggests that Batu Caves successfully meets or exceeds visitor expectations in many areas. Consequently, these insights inform efforts to maintain and enhance aspects that contribute to visitor satisfaction, ensuring continued positive evaluations.

The characteristics or types of visitors to Batu Caves are predominantly couples, families, friends, and solo travelers. This diverse range of visitor types reflects the broad appeal of Batu Caves, catering to various social groups and travel preferences. Each category of visitors brings unique expectations and experiences, contributing to the multifaceted allure of the destination. Understanding the dominance of these visitor types provides valuable insights for tailoring services and amenities to meet specific needs. Consequently, recognizing and addressing the preferences of these key visitor groups enhances overall satisfaction and promotes Batu Caves as an inclusive and attractive destination for all types of travelers.

Based on understanding the business context, proceeding to the data preparation stage is essential. This phase involves cleaning, transforming, and organizing the data to ensure its suitability for analysis. Proper data preparation is crucial as it directly impacts the accuracy and reliability of subsequent analytical results. Potential issues such as inconsistencies, missing values, and irrelevant information are addressed by meticulously preparing the data. Consequently, a well-prepared dataset forms the foundation for robust analysis, enabling meaningful insights and informed decision-making to enhance business outcomes.

### 2.2.2 Data Understanding

In the data understanding phase, it is essential to identify frequently used words and emojis. This step involves analyzing the textual data to uncover common expressions and symbols that reflect visitor sentiments and experiences. Recognizing these elements provides insights into the themes and emotions prevalent in visitor reviews, contributing to a deeper understanding of perceptions. Analyzing frequently used words and emojis is a foundation for more advanced sentiment analysis and text-mining techniques. Consequently, this process enhances the overall comprehension of visitor feedback, informing more targeted and effective strategies for improving the visitor experience.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**


**Figure 5.** Frequently used Words (Words Cloud)

Figure 5 shows the frequently used words in the first and second vlog videos. Based on the identification of the most frequently occurring words in the first video, it is evident that terms such as "nice" (38 times), "videos" (37 times), and "watching" (34 times) dominate the comments. Additionally, words like "love" (34 times), "visit" (33 times), and "beautiful" (31 times) are also prevalent, reflecting positive viewer sentiments. This lexical analysis suggests that viewers have a highly favorable perception of the video content, often expressing gratitude and admiration.

In addition, the most frequently occurring words in the second video, it is evident that terms such as "temple" (38 times), "like" (30 times), and "place" (27 times) dominate the comments. Additionally, words like "beautiful" (26 times), "video" (26 times), and "Batu" (25 times) are also prevalent, reflecting positive viewer sentiments and specific references to the destination. This lexical analysis suggests that viewers have a highly favorable perception of the video content, often emphasizing the beauty and significance of the location.

### 2.2.3 Modeling

In the modeling phase, VADER and TextBlob are compared based on the distribution of polarity values. This comparison involves analyzing how each tool classifies the sentiment polarity of the text data, providing insights into respective performance and accuracy. The distribution of polarity values highlights the differences in sentiment detection between the two methods, revealing strengths and potential biases. Examining these distributions allows for determining which tool offers a more reliable and nuanced sentiment analysis. Consequently, this comparative analysis informs the selection of the most appropriate sentiment analysis tool for the dataset, enhancing the overall robustness of the study's findings.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

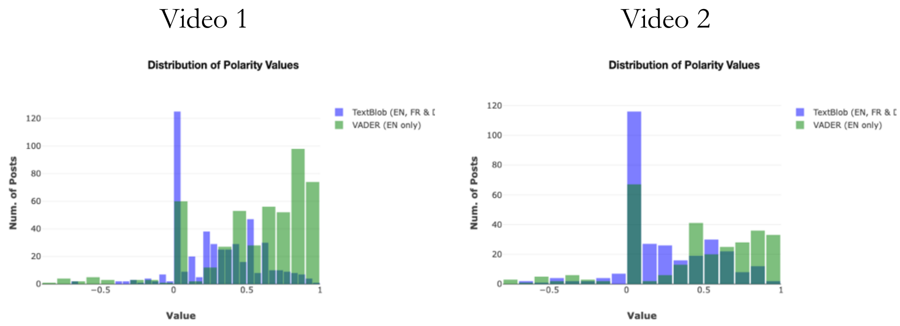



**Figure 6.** Distribution of Polarity Value between TextBlob and Vader of First and Second Video.

Figure 6 shows the distribution of the polarity value of the Vader and TextBlob Models. Based on the context of the first video, it is observed that VADER and TextBlob agree on the categorization of 364 (77.12%) out of 472 English language posts. This agreement level, reflected by Cohen's kappa statistic of 0.390, is considered fair. Specifically, both libraries concur on 12 (3.30%) posts with negative sentiments (polarity scores <= -0.05), 36 (9.89%) posts with neutral sentiments (polarity scores between -0.05 and 0.05), and 316 (86.81%) posts with positive sentiments (polarity scores >= 0.05).

In addition, Based on the context of the second video, it is observed that VADER and TextBlob agree on the categorization of 224 (76.45%) out of 293 English language posts. This agreement level is considered moderate, reflected by Cohen's kappa statistic of 0.519. Specifically, both libraries concur on 10 (4.46%) posts with negative sentiments (polarity scores <= -0.05), 51 (22.77%) posts with neutral sentiments (polarity scores between -0.05 and 0.05), and 163 (72.77%) posts with positive sentiments (polarity scores >= 0.05). This concordance underscores a reliable level of sentiment categorization, suggesting moderate consistency between the two tools.

### 2.2.4 Evaluation

The text data from the TripAdvisor platform is evaluated by analyzing the confusion matrix values, specifically accuracy, precision, recall, F-measure, and AUC. This comprehensive approach ensures a detailed assessment of the classification model's performance, providing insights into its strengths and weaknesses. Precision and recall measure the model's ability to correctly identify positive instances, while the F-measure balances these two metrics, offering a single performance score. The AUC further evaluates the model's overall ability to

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

distinguish between classes. Consequently, this rigorous evaluation framework enhances the reliability and validity of the sentiment analysis results.

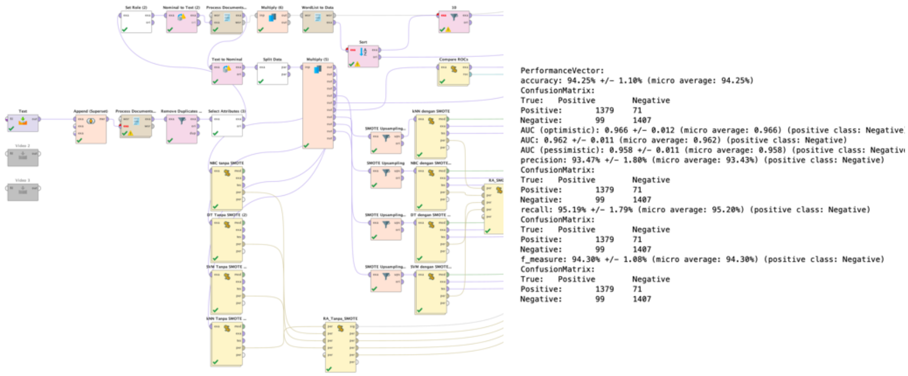



**Figure 7.** Evaluation of model Performance (Rapidminer)

Figure 7 shows the evaluation of model performance using Rapidminer. The algorithm with the best performance will be recommended as the optimal model for sentiment data classification. This selection process involves a rigorous comparison of various algorithms based on key performance metrics such as accuracy, precision, recall, and F-measure. By evaluating these metrics, the most effective algorithm for correctly identifying and categorizing sentiments will be identified. This approach ensures that the chosen model meets and exceeds the required performance standards. Consequently, recommending the best-performing algorithm enhances the reliability and accuracy of sentiment analysis, providing valuable insights for decision-making processes.

Subsequently, the review data from vlog content on YouTube is evaluated based on toxicity scores and the distribution of polarity values, categorized as poor, moderate, or fair. This evaluation involves analyzing the sentiment polarity to determine the overall tone of the comments, while toxicity scores measure the presence of harmful or abusive language. The analysis provides a nuanced understanding of viewer sentiment by classifying the polarity distribution into distinct categories. This dual approach ensures a comprehensive assessment of the content's reception. Consequently, these insights inform strategies for improving content quality and fostering a positive viewer community.

### 2.2.5 Deployment

In the deployment phase, the outputs from the data processing of TripAdvisor and YouTube reviews serve as a foundation for policy-making regarding the development of facilities and infrastructure related to the Batu Caves tourist

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

destination. These insights provide a data-driven basis for understanding visitor preferences and areas needing improvement. Leveraging this information enables targeted enhancements to boost visitor satisfaction and overall experience significantly. Consequently, informed policy decisions based on these analytics drive effective and sustainable development initiatives, ensuring that Batu Cave continues to attract and delight tourists.

Based on the findings of this study, policymakers align visitor preferences as the foundation for developing tourism programs at Batu Cave in Kuala Lumpur, Malaysia. The data-driven insights offer a clear understanding of visitor expectations and satisfaction levels, guiding the formulation of targeted strategies. By incorporating these preferences into development plans, it is possible to enhance the overall visitor experience and ensure sustainable tourism growth. Consequently, leveraging this research enables informed decision-making that effectively addresses visitor needs and promotes the long-term success of Batu Cave as a premier tourist destination.

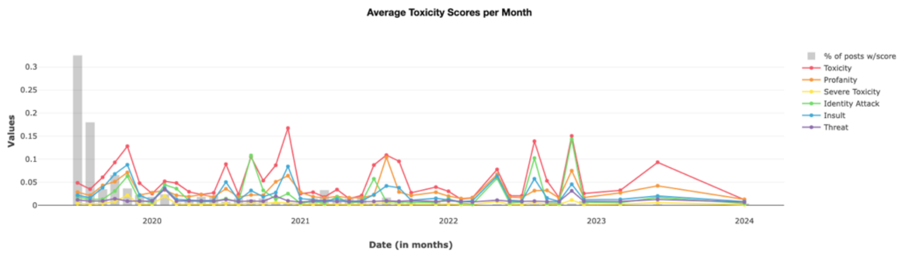
## 3. RESULTS AND DISCUSSION

### 3.1 Toxicity Score and Sentiment Model Evaluation based on Vlog Content of Batu Cave in Youtube Platform

Identifying toxicity scores for the first video, based on 550 posts out of 636 using the Perspective API, reveals significant findings. The toxicity score ranges from 0.05055 to 0.89882, while severe toxicity scores vary between 0.00441 and 0.35368. Additionally, identity attack scores range from 0.01910 to 0.62276, insult scores span from 0.02557 to 0.85022, profanity scores range from 0.03010 to 0.64460, and threat scores vary between 0.01080 and 0.34804. These metrics highlight the varying degrees of toxic content present in the posts. Consequently, these insights guide the implementation of measures to mitigate such content, fostering a healthier and more respectful online environment.

Figure 8 shows the toxicity score of the first and second videos. Identifying toxicity scores for the second video, based on 346 posts out of 433 using the Perspective API, reveals notable results. The toxicity scores range from 0.09896 to 0.82048, while severe toxicity scores vary between 0.01446 and 0.69417. Additionally, identity attack scores span from 0.04810 to 0.74265, insult scores range from 0.05347 to 0.76573, profanity scores vary between 0.05133 and 0.77964, and threat scores range from 0.01796 to 0.44185. These metrics highlight the presence and degree of toxic content in the posts.

First Video




Second Video

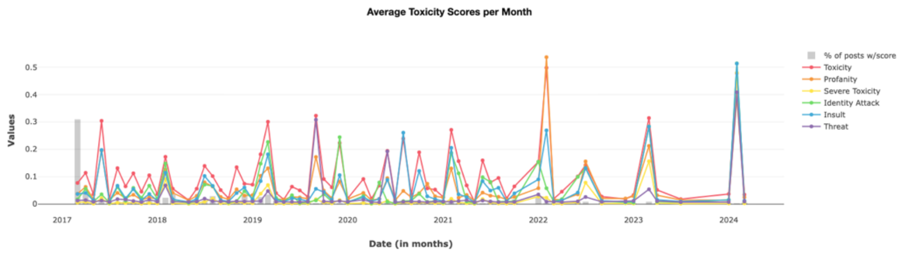



**Figure 8.** Toxicity Score of First and Second Videos

Subsequently, the sentiment analysis results of 494 out of 636 posts for the first video reveal distinct patterns. According to VADER (English/EN), out of 484 posts, 21 (4.34%) exhibit negative sentiment, 57 (11.78%) show neutral sentiment, and 406 (83.88%) reflect positive sentiment. TextBlob (English/EN) classifies 484 posts with 24 (4.96%) negative, 120 (24.79%) neutral, and 340 (70.25%) positive sentiments. TextBlob (French/FR) evaluates two neutral posts, while TextBlob (German/DE) assesses four posts, all showing neutral sentiment. These results highlight the overall positive reception of the video content, with VADER indicating a higher proportion of positive sentiments than TextBlob.

Figure 9 shows the sentiment classification results of the first and second videos. Subsequently, the sentiment analysis results of 311 out of 433 posts for the second video reveal distinct patterns. According to VADER (English/EN), out of 293 posts, 22 (7.51%) exhibit negative sentiment, 65 (22.18%) show neutral sentiment, and 206 (70.31%) reflect positive sentiment. TextBlob (English/EN) classifies 293 posts with 21 (7.17%) negative, 90 (30.72%) neutral, and 182 (62.12%) positive sentiments. TextBlob (French/FR) evaluates 15 neutral posts, while TextBlob (German/DE) assesses two posts, all showing neutral sentiment. These results highlight the overall positive reception of the video content, with VADER indicating a higher proportion of positive sentiments than TextBlob.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Frist Video

| | # of Posts | Negative Sentiment [-1..-0.05] | Neutral Sentiment (-0.05..0.05) | Positive Sentiment [0.05..1] |
|---|---|---|---|---|
| VADER (English/EN) | 484 | 21 (4.34%) | 57 (11.78%) | 406 (83.88%) |
| TextBlob (English/EN) | 484 | 24 (4.96%) | 120 (24.79%) | 340 (70.25%) |
| TextBlob (French/FR) | 2 | 0 (0.00%) | 2 (100.00%) | 0 (0.00%) |
| TextBlob (German/DE) | 4 | 0 (0.00%) | 4 (100.00%) | 0 (0.00%) |

Second Video

| | # of Posts | Negative Sentiment [-1..-0.05] | Neutral Sentiment (-0.05..0.05) | Positive Sentiment [0.05..1] |
|---|---|---|---|---|
| VADER (English/EN) | 293 | 22 (7.51%) | 65 (22.18%) | 206 (70.31%) |
| TextBlob (English/EN) | 293 | 21 (7.17%) | 90 (30.72%) | 182 (62.12%) |
| TextBlob (French/FR) | 15 | 0 (0.00%) | 15 (100.00%) | 0 (0.00%) |
| TextBlob (German/DE) | 2 | 0 (0.00%) | 2 (100.00%) | 0 (0.00%) |

**Figure 9.** Sentiment Classification of First and Second Video

### 3.2 Sentiment Classification of Batu Cave Reviews on TripAdvisor Using Support Vector Machine Algorithm

Based on the text data analysis from TripAdvisor reviews about Batu Cave, it is evident that certain words frequently appear in visitor responses. Words such as "captivating," "behavior," "application," and "appearance" are commonly mentioned, reflecting key aspects of visitor experiences and observations. This frequent usage indicates that these elements significantly influence visitors' perceptions and satisfaction. Analyzing these standard terms provides valuable insights into the factors that shape visitor opinions, thereby informing strategies to enhance the tourist experience at Batu Cave.

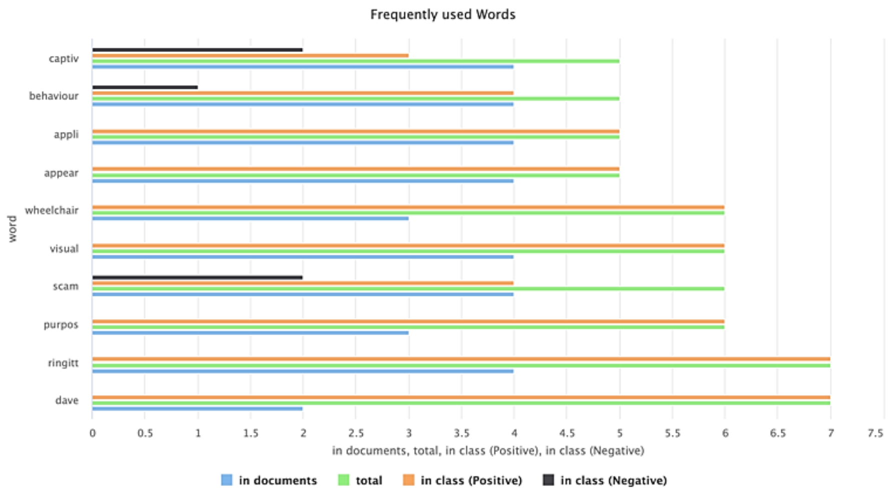



**Figure 10.** Frequently used Words of TripAdvisor Dataset

Figure 10 shows the frequently used words based on the TripAdvisor dataset. Nevertheless, words such as "wheelchair," "scam," and "behavior" indicate that visitors with disabilities are unable to enjoy visits due to a lack of supporting facilities entirely. Additionally, the term "scam" suggests the prevalence of fraudulent activities. Furthermore, "behavior" reflects the significant impact of the conduct of both visitors and hosts on the overall visitor experience. These insights highlight critical areas for improvement to ensure a more inclusive and secure environment for all visitors.

In addition, visitor perceptions, motives, and impressions when traveling to Batu Cave were discerned. Words such as "captivating," "beautiful," and "temple" suggest a strong appreciation for the site's aesthetic and cultural significance. Additionally, terms like "steps" and "climb" highlight the physical aspects of the visit, indicating that the journey itself is a memorable part of the experience. These lexical patterns reveal that visitors are motivated by both the visual appeal and the unique challenges of the destination.


**Figure 11.** Support Vector Machine Performance

Figure 11 shows the SVM performance. The performance of the SVM model is notably high, as indicated by a PerformanceVector accuracy of 94.25% ± 1.10% (micro average: 94.25%). The ConfusionMatrix results reveal that the model correctly classified 1379 optimistic and 1407 negative cases, with 71 false positives and 99 false negatives. The AUC (Area Under the Curve) scores further underscore the model's robustness, with optimistic, standard, and pessimistic AUC values of 0.966 ± 0.012, 0.962 ± 0.011, and 0.958 ± 0.011, respectively, for the negative class. Precision is recorded at 93.47% ± 1.80% (micro average: 93.43%), while recall stands at 95.19% ± 1.79% (micro average: 95.20%). The f_measure, which balances precision and recall, is 94.30% ± 1.08% (micro average: 94.30%). These metrics collectively highlight the SVM model's efficacy in accurately classifying sentiments with high reliability and minimal error.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

The Area Under the Curve (AUC) scores further underscore the model's robustness, with optimistic, standard, and pessimistic AUC values of 0.966 ± 0.012, 0.962 ± 0.011, and 0.958 ± 0.011, respectively, for the negative class. These values indicate a high level of accuracy in distinguishing between positive and negative instances under various conditions. The minimal variance in these scores highlights the model's stability and reliability in sentiment classification. Such high AUC values demonstrate the model's exceptional performance, suggesting its effectiveness for practical applications in sentiment analysis. Consequently, these metrics validate the model's suitability for deployment in real-world scenarios, ensuring consistent and reliable results.

### 3.3 Discussion

Sentiment analysis is closely linked to tourist behavior and preferences, facilitating destination managers in understanding visitor characteristics. By examining the sentiments expressed in reviews and feedback, it is possible to gain insights into what tourists appreciate or find lacking in the experiences [51]. This analysis highlights general visitor satisfaction and pinpoints specific areas that require attention or improvement [52]. Consequently, leveraging sentiment analysis allows for a more targeted approach to enhancing the visitor experience, ensuring that tourism strategies are aligned with actual visitor needs and expectations. This method ultimately aids in creating more engaging and satisfactory tourist destinations.

The outcomes from calculating toxicity scores on vlog videos about Batu Cave serve as a foundation for policy-making to enhance facilities and services to meet tourist needs. By identifying toxicity levels in visitor comments, specific issues such as the need for improved safety, better visitor conduct, and enhanced accessibility [53]. This data-driven approach ensures that any interventions are directly aligned with the concerns and expectations of tourists [54]. Consequently, leveraging toxicity score analysis enables the development of targeted strategies to improve the overall visitor experience, ensuring that facilities and services are comprehensive and practical.

Subsequently, the sentiment classification results based on TripAdvisor review data utilized to optimize the marketing of Batu Cave as a destination. Analyzing visitor sentiments provides valuable insights into the destination's most appreciated aspects, such as cultural significance and natural beauty [55]. Additionally, identifying areas of concern or dissatisfaction enables targeted improvements to be highlighted in the marketing campaign f[56]. This strategic use of sentiment data ensures that marketing efforts resonate more effectively with potential visitors, emphasizing the strengths of Batu Cave while addressing any perceived weaknesses. Consequently, leveraging sentiment classification enhances the precision and impact of marketing strategies, attracting a broader and more satisfied visitor base.

Thus, the findings of this study contribute both theoretically and practically. Theoretically, the research advances the understanding of sentiment analysis methodologies and applications in tourism studies, offering insights into visitor behavior and preferences. Practically, the results provide actionable data for destination managers, enabling them to tailor services and marketing strategies to meet tourist expectations better. This dual contribution enriches academic discourse and enhances the operational effectiveness of tourism management. Consequently, the study bridges the gap between theory and practice, ensuring its relevance and applicability in the real world.

## 4. CONCLUSION

The research conclusion highlights the significant findings and implications of the study utilizing the CRISP-DM framework. The analysis demonstrates that sentiment analysis tools like VADER and TextBlob provide valuable insights into visitor perceptions and experiences at Batu Cave. The high accuracy (94.25%) and AUC scores (optimistic: 0.966, standard: 0.962, pessimistic: 0.958) of the SVM model further validate its effectiveness in sentiment classification, offering a robust method for understanding tourist feedback. Additionally, evaluating toxicity scores using the Perspective API reveals critical areas for improvement in managing visitor interactions, with toxicity levels ranging from 0.05055 to 0.89882. Following the CRISP-DM methodology, the study systematically processed and analyzed the data, ensuring comprehensive and reliable results. These findings underscore the importance of leveraging advanced analytical techniques to enhance destination management and improve visitor satisfaction. Consequently, the study provides a solid foundation for informed decision-making and strategic planning in the tourism sector, ensuring the sustainable development of tourist destinations.

## REFERENCES

[1] H. Xu, L. T. O. Cheung, J. Lovett, X. Duan, Q. Pei, and D. Liang, "Understanding the influence of user-generated content on tourist loyalty behavior in a cultural World Heritage Site," *Tour. Recreat. Res.*, vol. 48, no. 2, pp. 173–187, 2023, doi: 10.1080/02508281.2021.1913022.

[2] S. M. Braimah, E. N. A. Solomon, and R. E. Hinson, "Tourists satisfaction in destination selection determinants and revisit intentions; perspectives from Ghana," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2318864.

[3] I. Leri and P. Theodoridis, "How do personality traits affect visitor's experience, emotional stimulation and behaviour? The case of wine tourism," *Tour. Rev.*, vol. 76, no. 5, pp. 1013–1049, Jan. 2021, doi: 10.1108/TR-05-2019-0148.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[4] A. H. Fansurya *et al.*, "Tourist's Length of stay: the perspective of flow experience theory," *Cogent Bus. Manag.*, vol. 11, no. 1, p., 2024, doi: 10.1080/23311975.2024.2310258.
[5] I. H. Blomstervik, N. K. Prebensen, A. C. Campos, and P. Pinto, "Novelty in tourism experiences: the influence of physical staging and human interaction on behavioural intentions," *Curr. Issues Tour.*, vol. 24, no. 20, pp. 2921–2938, 2021, doi: 10.1080/13683500.2020.1854197.
[6] S. Terasaki, T. Hara, and J. Ikegami, "Mediating role of the country image in enhancing memorable experiences and revisits: an Analysis of U.S. tourists in Japan," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–13, 2023, doi: 10.1080/02508281.2023.2185733.
[7] S. Wang, Y. Huang, and T. Li, "Understanding visitor flow and behaviour in developing tourism-service-oriented villages by space syntax methodologies: a case study of Tabian Rural Section of Qingshan Village, Hangzhou," *J. Asian Archit. Build. Eng.*, vol. 00, no. 00, pp. 1–20, 2024, doi: 10.1080/13467581.2024.2349737.
[8] E. Sthapit, B. Garrod, D. N. Coudounaris, P. Björk, E. Erul, and H. Song, "Antecedents and outcomes of memorable wildlife tourism experiences," *J. Ecotourism*, 2023, doi: 10.1080/14724049.2023.2272063.
[9] A. G. Asmelash and S. Kumar, "Tourist satisfaction-loyalty Nexus in Tigrai, Ethiopia: Implication for sustainable tourism development," *Cogent Bus. Manag.*, vol. 7, no. 1, 2020, doi: 10.1080/23311975.2020.1836750.
[10] S. Thipsingh, W. A. Srisathan, S. Wongsaichia, C. Ketkaew, P. Naruetharadhol, and L. Hengboriboon, "Social and sustainable determinants of the tourist satisfaction and temporal revisit intention: A case of Yogyakarta, Indonesia," *Cogent Soc. Sci.*, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2022.2068269.
[11] K. J. Dsouza, A. Shetty, P. Damodar, A. D. Shetty, and T. K. Dinesh, "The assessment of locavorism through the lens of agritourism: the pursuit of tourist's ethereal experience," *Econ. Res. Istraz.* , vol. 36, no. 2, p., 2023, doi: 10.1080/1331677X.2022.2147566.
[12] M. Saad, N. A. Husin, S. N. H. Rosman, R. Husain, and L. Abdul Latif @ Bapooe, "The nexus of community perception on turning idyllic Panchor town into a future river tourism destination," *Cogent Soc. Sci.*, vol. 9, no. 2, 2023, doi: 10.1080/23311886.2023.2253021.
[13] K. Xiang, C. Xu, and J. Wang, "Understanding the relationship between tourists' consumption behavior and their consumption substitution willingness under unusual environment," *Psychol. Res. Behav. Manag.*, vol. 14, pp. 483–500, 2021, doi: 10.2147/PRBM.S303239.
[14] R. Singh, M. A. Mir, and A. A. Nazki, "Evaluation of tourist behavior towards traditional food consumption: validation of extended Theory of Planned Behaviour," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2023.2298893.
[15] A. Ghorbani, H. Mousazadeh, A. Golafshan, F. Akbarzadeh Almani, and L. D. Dávid, "Tourists-generated photographs in online media and tourism destination choice: The case of Shiraz metropolis in Iran," *Cogent Soc. Sci.*, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2225336.
[16] S. Bingöl, "From escape to seeking: understanding drug tourists," *J. Tour. Cult. Chang.*, vol. 20, no. 4, pp. 583–599, 2022, doi: 10.1080/14766825.2021.1960853.
[17] K. T. Phan, S. H. Chen, J. M. Lee, and C. Van Pham, "Expenditure decisions of

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

international tourists to Taiwan: application of the Heckman two-stage approach," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2318870.
[18] A. Abror *et al.*, "Antecedents of Muslim tourist loyalty: The role of Islamic religiosity and tourist value co-creation," *Cogent Bus. Manag.*, vol. 10, no. 2, 2023, doi: 10.1080/23311975.2023.2247871.
[19] L. W. Liu, C. C. Wang, P. Pahrudin, A. F. Royanow, C. Lu, and I. Rahadi, "Does virtual tourism influence tourist visit intention on actual attraction? A study from tourist behavior in Indonesia," *Cogent Soc. Sci.*, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2240052.
[20] M. A. I. Gazi, M. A. Islam, A. Al Masud, A. R. bin S. Senathirajah, S. Biswas, and R. A. Shuvro, "The moderating impacts of COVID-19 fear on hotel service quality and tourist satisfaction: Evidence from a developing country," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2331079.
[21] B. Pikkemaat, B. F. Bichler, and M. Peters, "Exploring the crowding-satisfaction relationship of skiers: the role of social behavior and experiences," *J. Travel Tour. Mark.*, vol. 37, no. 8–9, pp. 902–916, 2020, doi: 10.1080/10548408.2020.1763229.
[22] T. Nguyen Huu, H. Nguyen Ngoc, L. Nguyen Dai, D. Nguyen Thi Thu, L. N. Truc, and L. Nguyen Trong, "Effect of tourist satisfaction on revisit intention in Can Tho City, Vietnam," *Cogent Bus. Manag.*, vol. 11, no. 1, p., 2024, doi: 10.1080/23311975.2024.2322779.
[23] S. Seyfi, S. Kuhzady, R. Rastegar, T. Vo-Thanh, and M. Zaman, "Exploring the dynamics of tourist travel intention before and during the COVID-19 pandemic: a scoping review," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–14, 2024, doi: 10.1080/02508281.2024.2330236.
[24] R. Martínez Suárez, J. A. Castañeda García, and M. Á. Rodríguez Molina, "Identifying tourist profiles to reduce overtourism: the case of a cultural destination," *Int. J. Tour. Cities*, vol. 7, no. 4, pp. 962–985, 2021, doi: 10.1108/IJTC-08-2020-0153.
[25] A. Galiano-Coronil, S. Blanco-Moreno, L. B. Tobar-Pesantez, and G. A. Gutiérrez-Montoya, "Social media impact of tourism managers: a decision tree approach in happiness, social marketing and sustainability," *J. Manag. Dev.*, vol. 42, no. 6, pp. 436–457, 2023, doi: 10.1108/JMD-04-2023-0131.
[26] S. H. Laaksonen and P. Varga, "Assessing the impact of selfie-taking tourists on local tour guides in the Chernobyl Exclusion Zone: a netnographic analysis of a dark tourism location," *J. Herit. Tour.*, 2023, doi: 10.1080/1743873X.2023.2292147.
[27] A. Kokkhangplu, "A new generation behavior intention to visit tourism destinations in perspective of post-COVID-19," *Cogent Bus. Manag.*, vol. 11, no. 1, p., 2024, doi: 10.1080/23311975.2024.2317461.
[28] J. Räikkönen, M. Grénman, H. Rouhiainen, A. Honkanen, and I. E. Sääksjärvi, "Conceptualizing nature-based science tourism: a case study of Seili Island, Finland," *J. Sustain. Tour.*, vol. 31, no. 5, pp. 1214–1232, 2023, doi: 10.1080/09669582.2021.1948553.
[29] I. W. B. Suryawan, N. W. Utami, and K. Q. Fredlina, "Analisis Sentimen Review Wisatawan Pada Objek Wisata Ubud Menggunakan Algoritma Support Vector Machine," *JINTEKS (Jurnal Inform. Teknol. dan Sains)*, vol. 5, no. 1, pp. 133–140, 2023.
[30] A. A. Arifiyanti, M. Fuad, P. Fikri, and B. Utomo, "Analisis Sentimen Ulasan Pengunjung Objek Wisata Gunung Bromo pada Situs Tripadvisor," *Explor. J. Sist.*

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

*Inf. dan Telemat.*, vol. 13, no. 1, pp. 32–37, 2022.
[31] Y. A. Singgalen, "Social Network Analysis and Sentiment Classification of Extended Reality Product Content," *J. Tek. Inform. C.I.T Medicom*, vol. 16, no. 1, pp. 24–34, 2024.
[32] M. B. Duignan and I. Pappalepore, "How do Olympic cities strategically leverage New Urban Tourism? Evidence from Tokyo," *Tour. Geogr.*, vol. 25, no. 2–3, pp. 425–449, 2023, doi: 10.1080/14616688.2021.1895296.
[33] Z. Gao *et al.*, "Exploring tourist spatiotemporal behavior differences and tourism infrastructure supply–demand pattern fusing social media and nighttime light remote sensing data," *Int. J. Digit. Earth*, vol. 17, no. 1, pp. 1–22, 2024, doi: 10.1080/17538947.2024.2310723.
[34] N. Sharma, "Dark tourism and moral disengagement in liminal spaces," *Tour. Geogr.*, vol. 22, no. 2, pp. 273–297, 2020, doi: 10.1080/14616688.2020.1713877.
[35] O. M. Stoleriu, B. C. Ibănescu, C. C. Stoleriu, and C. Lupu, "Strengthening Dracula tourism brand through cartographic approaches," *J. Maps*, vol. 18, no. 1, pp. 61–69, 2022, doi: 10.1080/17445647.2022.2071647.
[36] K. K. Smørvik, "Contagious behaviour or not? Tourists' skills and practices in a Norwegian church," *J. Herit. Tour.*, vol. 18, no. 6, pp. 756–767, 2023, doi: 10.1080/1743873X.2023.2250470.
[37] B. McKercher, B. Prideaux, and M. Thompson, "The relationship between accommodation type and tourists' in-destination behaviour," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–10, 2023, doi: 10.1080/02508281.2023.2221070.
[38] K. Berhanu and S. Raj, "The role of social media marketing in Ethiopian tourism and hospitality organizations: Applying the unified theory of acceptance and use of technology model," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2318866.
[39] R. Iloranta and R. Komppula, "Service providers' perspective on the luxury tourist experience as a product," *Scand. J. Hosp. Tour.*, vol. 22, no. 1, pp. 39–57, 2022, doi: 10.1080/15022250.2021.1946845.
[40] Y. Jeaheng *et al.*, "Influence of Thai Street Food Quality, Price, and Involvement on Traveler Behavioral Intention: Exploring Cultural Difference (Eastern versus Western)," *Psychol. Res. Behav. Manag.*, vol. 16, pp. 223–240, 2023, doi: 10.2147/PRBM.S371806.
[41] W. H. Liang, "Cultural friction during intercultural service encounters with Chinese tourists: perspectives from hotel employees in Australia," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2318868.
[42] X. Wujie, "The impact of geopolitical risks and international relations on inbound tourism——evidence from China and key source countries," *Cogent Soc. Sci.*, vol. 9, no. 2, 2023, doi: 10.1080/23311886.2023.2285244.
[43] M. Šerić, J. Mikulić, and Đ. Ozretić Došen, "Understanding prevention measures and tourist behavior in Croatia during the COVID-19 pandemic. A mixed-method approach," *Econ. Res. Istraz.* , vol. 36, no. 2, p., 2023, doi: 10.1080/1331677X.2022.2135556.
[44] S. Ding, R. Zhang, Y. Liu, P. Lu, and M. Liu, "Visitor crowding at World Heritage Sites based on tourist spatial-temporal distribution: a case study of the Master-of-Nets Garden, China," *J. Herit. Tour.*, vol. 18, no. 5, pp. 632–657, 2023, doi: 10.1080/1743873X.2023.2214680.
[45] U. Zaman, F. Bayrakdaroglu, M. Gohar Qureshi, M. Aktan, and S. Nawaz, "Every

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

storm will pass: Examining expat's host country-destination image, cultural intelligence and renewed destination loyalty in COVID-19 tourism," *Cogent Bus. Manag.*, vol. 8, no. 1, 2021, doi: 10.1080/23311975.2021.1969631.

[46] L. Chang, X. Huang, and M. Meng, "Study on tourist's loyalty of Zhinan Village in the view of tourism landscape," *Cogent Soc. Sci.*, vol. 7, no. 1, 2021, doi: 10.1080/23311886.2021.1997403.

[47] F. Bazzanella, M. Schnitzer, M. Peters, and B. F. Bichler, "The role of sports events in developing tourism destinations: a systematized review and future research agenda," *J. Sport Tour.*, vol. 27, no. 2, pp. 77–109, 2023, doi: 10.1080/14775085.2023.2186925.

[48] F. Quoquab, J. Mohammad, and A. M. Mohd Sobri, "Psychological engagement drives brand loyalty: evidence from Malaysian ecotourism destinations," *J. Prod. Brand Manag.*, vol. 30, no. 1, pp. 132–147, Jan. 2021, doi: 10.1108/JPBM-09-2019-2558.

[49] E. Kim, S. Kim, and Y. Jeong, "The 18th FINA world masters championships and destination loyalty," *Asia Pacific J. Mark. Logist.*, vol. 36, no. 3, pp. 622–640, Jan. 2024, doi: 10.1108/APJML-04-2023-0294.

[50] Y. A. Singgalen, "Sentiment classification of coral reef 101 content using decision tree algorithm through CRISP-DM," *Int. J. Basic Appl. Sci.*, vol. 12, no. 3, pp. 121–130, 2023.

[51] X. Zhan, F. bin Guo, and E. Roberts, "Image-based research methods for mapping tourist behaviour: smart photos," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–7, 2023, doi: 10.1080/02508281.2023.2196487.

[52] T. D. Quang, N. M. P. Tran, E. Sthapit, and B. Garrod, "Exploring Guests' Satisfaction and Dissatisfaction with Homestay Experiences: A Netnographic Study of a Rural Tourism Destination in Vietnam," *Int. J. Hosp. Tour. Adm.*, vol. 00, no. 00, pp. 1–25, 2024, doi: 10.1080/15256480.2024.2350005.

[53] H. Shen, N. F. Aziz, M. Huang, L. Yu, and Z. Liu, "Tourist perceptions of landscape in Chinese traditional villages: analysis based on online data," *J. Tour. Cult. Chang.*, vol. 22, no. 2, pp. 232–251, 2024, doi: 10.1080/14766825.2023.2240297.

[54] K. Puh and M. Bagić Babac, "Predicting sentiment and rating of tourist reviews using machine learning," *J. Hosp. Tour. Insights*, vol. 6, no. 3, pp. 1188–1204, Jan. 2023, doi: 10.1108/JHTI-02-2022-0078.

[55] M. N. Islam, "Visitors' Sentiment in Archeological Heritage Sites: A Study Based on TripAdvisor Reviews," *Int. J. Hosp. Tour. Adm.*, vol. 00, no. 00, pp. 1–22, 2023, doi: 10.1080/15256480.2023.2242343.

[56] A. Leiras and C. Eusébio, "Perceived image of accessible tourism destinations: a data mining analysis of Google Maps reviews," *Curr. Issues Tour.*, vol. 0, no. 0, pp. 1–19, 2023, doi: 10.1080/13683500.2023.2230338.