KLIK: Kajian Ilmiah Informatika dan Komputer
ISSN 2723-3898 (Media Online)
Vol 4, No 6, June 2024, Hal 2992-3004
DOI 10.30865/klik.v4i6.1947
https://djournals.com/klik

Understanding Digital Engagement through Sentiment Analysis of Tourism Destination through Travel Vlog Reviews

Yerik Afrianto Singgalen

Faculty of Business Administration and Communication, Tourism Department, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: yerik.afrianto@atmajaya.ac.id
Email Penulis Korespondensi: yerik.afrianto@atmajaya.ac.id

Abstract−This research employs the CRISP-DM framework to analyze digital engagement through travel vlog content, explicitly focusing on vlogs about Gili Trawangan. The study systematically follows the CRISP-DM phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Utilizing the VADER sentiment analysis model and the SVM algorithm with SMOTE, the research achieves a high level of accuracy in sentiment classification, with the SVM model demonstrating an accuracy of 88.57% +/- 5.11% and a precision of 90.95% +/- 5.09%. Analysis of 442 cleaned and labeled data points reveals a strong dominance of positive sentiments, with 62.61% in the first video and 84.25% in the second video. These findings underscore the effectiveness of travel vlogs in engaging viewers and generating positive interactions as powerful tools for tourism marketing. The study concludes that the CRISP-DM framework is highly effective in facilitating comprehensive sentiment analysis and enhancing strategic tourism marketing initiatives.

Keywords: Tourist Behavior; Destination Marketing; Digital engagement; Travel Vlog; Tourism

1. INTRODUCTION

Analyzing tourist behavior is crucial to identifying market segments and examining preferences and trends for effective marketing strategies. Understanding the motivations and actions of tourists enables the development of targeted promotions and tailored services, enhancing overall satisfaction and loyalty [1]–[4]. By assessing demographic data, psychographic profiles, and consumption patterns, it is possible to predict emerging trends and adapt offerings to meet evolving demands [5]–[8]. Effective segmentation and preference analysis drive competitive advantage, enabling a precise alignment of marketing efforts with consumer expectations [9]–[12]. In conclusion, thoroughly examining tourist behavior provides invaluable insights for optimizing marketing strategies in the tourism industry.

The rise of travel vlogs across various tourist destinations significantly enhances destination marketing. Digital content uploaded to various media-sharing platforms, mainly social media, provides tourists with extensive information about the locations they visit [13]–[16]. These vlogs offer insights into local attractions, cultural experiences, and practical travel tips, influencing potential visitors' decisions and enriching travel planning [17]–[20]. Consequently, travel vlogs contribute to a broader and more engaged audience, fostering increased interest and visitation rates. In conclusion, the trend of travel vlogging plays an integral role in effectively promoting tourism destinations.

This study aims to analyze the content and discuss sentiment analysis results based on review data from travel vlog videos uploaded to YouTube. By examining the comments and feedback from viewers, valuable insights into user perceptions and attitudes towards specific tourist destinations are obtained [21], [22]. The analysis identifies key themes and sentiment trends, highlighting areas of positive reception and potential improvement [23]–[25]. It is argued that such insights are crucial for destination marketers to tailor strategies effectively [26]–[28]. In conclusion, the findings from this research provide a deeper understanding of user responses on the platform, enhancing the strategic promotion of tourism destinations.

This case study focuses on Gili Trawangan in Lombok, as reviewed in YouTube videos with the IDs s9ut2k72uwc and S7utYxLazMA. The analysis examines how these videos portray the destination's attractions and amenities and appeal to potential tourists. By assessing viewer engagement and sentiment in the comments section, insights are gained into public perceptions and expectations [29]–[32]. It is posited that such targeted analysis helps understand the effectiveness of digital content in promoting tourism destinations [33], [34]. In conclusion, this case study sheds light on the influential role of YouTube content in shaping the image and attractiveness of Gili Trawangan as a tourist destination.

The methodology employed in this research is the Cross-Industry Standard Process for Data Mining (CRISP-DM). This robust and structured approach encompasses six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Each phase ensures a comprehensive analysis, allowing for the systematic extraction of valuable insights from large datasets. Utilizing CRISP-DM facilitates a thorough and replicable process, enhancing the reliability and validity of the research findings. In conclusion, the application of CRISP-DM in this study provides a meticulous framework for data mining, contributing to the rigor and depth of the research methodology.

The theoretical and practical contributions of this study significantly advance tourism knowledge. Theoretically, the research introduces novel frameworks and models that deepen the understanding of tourist behavior, destination marketing, and digital engagement. Practically, the study provides actionable insights for tourism stakeholders, enabling them to develop more effective marketing strategies and enhance visitor experiences. It is posited that these contributions not only bridge gaps in existing literature but also drive innovation in tourism practices. In conclusion, the study's findings offer substantial value in academic discourse and real-world application, fostering growth and development within the tourism sector.

The limitation of this study lies in the context of the destination and video content, which is confined to the travel experiences in Gili Trawangan, as reviewed in YouTube videos with the IDs s9ut2k72uwc and S7utYxLazMA. This narrow focus restricts the generalizability of the findings to other destinations or types of video content. Additionally, the analysis is limited to the selected videos' subjective experiences and presentation styles, potentially introducing bias. Expanding the scope to include multiple destinations and diverse content formats could provide a more comprehensive understanding. In conclusion, while the current focus offers valuable insights into Gili Trawangan, broader studies are necessary for more generalized conclusions.

2. RESEARCH METHODOLOGY

2.1 Gap Analysis

Gap analysis is conducted to determine the extent of progress in studying Tourist Behavior, Destination Marketing, Digital Engagement, Travel Vlogs, and Tourism. This analysis helps identify the areas where current research is lacking and where future studies should be directed [35], [35]–[37]. By evaluating the existing literature and identifying discrepancies, the study uncovers critical insights into how these fields have evolved and where they intersect [38]–[41]. It is posited that addressing these gaps will significantly enhance the theoretical and practical understanding of these interconnected domains. In conclusion, gap analysis provides a comprehensive framework for advancing research in the multifaceted realm of tourism studies.

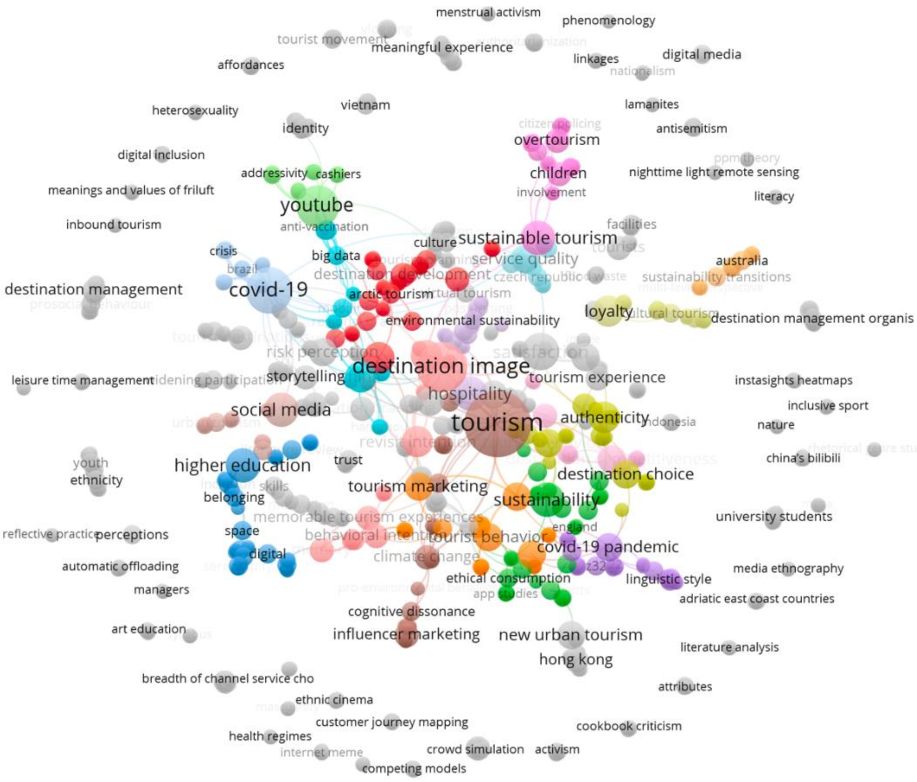


Figure 1. Gap Analysis using Vosviewer

Figure 1 shows the network, overlay, and density visualization. Considering the contextuality of destinations and travel vlog content, this research must understand aspects related to digital engagement, tourist behavior, and travel vlogs in the marketing media [42]–[45]. The study aims to explore how digital platforms influence tourist decision-making processes and behaviors by analyzing the interactions and feedback on travel vlogs. Additionally, it examines the effectiveness of travel vlogs in promoting tourist destinations, offering insights into marketing tools. It is argued that this understanding will enable more strategic utilization of digital content to attract and retain tourists. In conclusion, this research addresses critical intersections between digital engagement, tourist behavior, and travel vlogs, contributing valuable knowledge to tourism marketing.

Several studies on sentiment analysis remain focused on testing algorithms, neglecting the primary objective of understanding tourist behavior and the effectiveness of digital technologies for viewer engagement [46]–[50]. Sentiment analysis in tourism aims to decode tourist sentiments and preferences, offering insights into how digital tools influence experiences. This research, therefore, delves into the context of tourism destinations using a sentiment analysis approach guided by the CRISP-DM methodology. It is posited that this approach not only enhances the accuracy of sentiment analysis but also provides practical insights for improving digital engagement strategies. In conclusion, the study bridges the gap between algorithmic development and practical application in tourism sentiment analysis.

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

This study employs the CRISP-DM framework for data processing within the context of travel vlogs and the Gili Trawangan tourist destination in Nusa Tenggara Barat. The CRISP-DM methodology guides the systematic approach to data analysis, from business understanding to deployment, ensuring a structured and comprehensive examination of the digital content. The study aims to derive objective insights into tourist behavior and digital engagement by focusing on data understanding, modeling, and evaluation. It is posited that the application of CRISP-DM enhances the accuracy and relevance of the findings. In conclusion, this methodological approach provides a robust framework for analyzing travel vlog data and optimizing tourism marketing strategies.

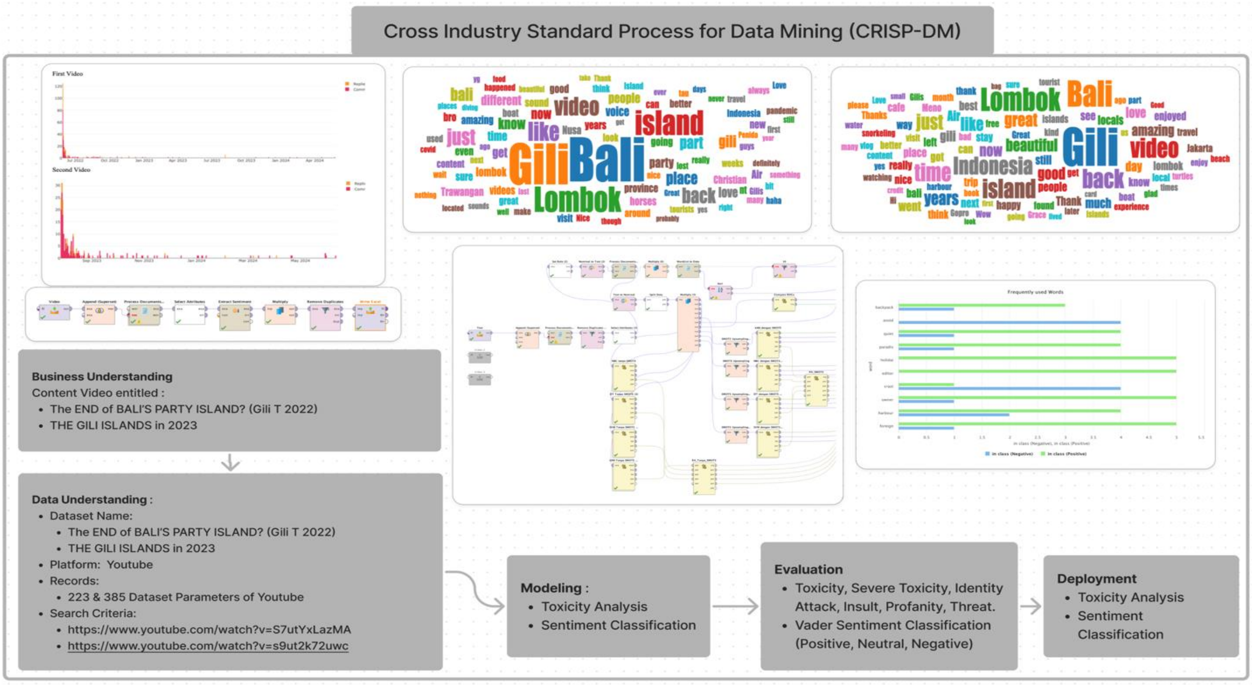


Figure 2. Implementation of CRISP-DM Framework

Figure 2 shows the implementation of the CRISP-DM framework. The utilization of the CRISP-DM framework is based on considerations of the business understanding and deployment phases, which yield outputs implemented to optimize business processes and enhance productivity across various fields, as demonstrated in case studies. The business understanding phase ensures a clear grasp of project objectives and requirements, aligning them with business goals. The deployment phase translates analytical results into actionable strategies, facilitating practical applications and measurable improvements. It is posited that CRISP-DM's structured approach enhances analytical precision and ensures practical relevance and scalability. In conclusion, the CRISP-DM framework effectively bridges the gap between data analysis and business optimization, demonstrating its versatility and efficacy.

Subsequently, the data understanding, modeling, and evaluation phases in the CRISP-DM framework strengthen data-driven arguments or recommendations, enhancing objectivity. The data understanding phase involves comprehensive data collection and exploration, ensuring the reliability and relevance of the dataset. The modeling phase applies various analytical techniques to uncover patterns and insights, while the evaluation phase assesses the model's performance and accuracy. This systematic approach ensures that conclusions drawn are well-supported by empirical evidence. In conclusion, these phases collectively contribute to the robustness and credibility of data-driven decisions, making them more objective and impactful.

2.2.1 Business Understanding

In the business understanding phase, it is essential to identify the data sources and characteristics that will be further processed. The data utilized in this study originates from the YouTube platform, specifically selected content about Gili Trawangan from the years 2022 (s9ut2k72uwc) and 2023 (S7utYxLazMA). Understanding the origin and nature of this data is crucial for ensuring its relevance and reliability in subsequent analyses. It is argued that this careful selection process enhances the validity of the study's findings. In conclusion, thoroughly comprehending the data sources and the attributes forms the foundation for effective data processing and analysis in this research.

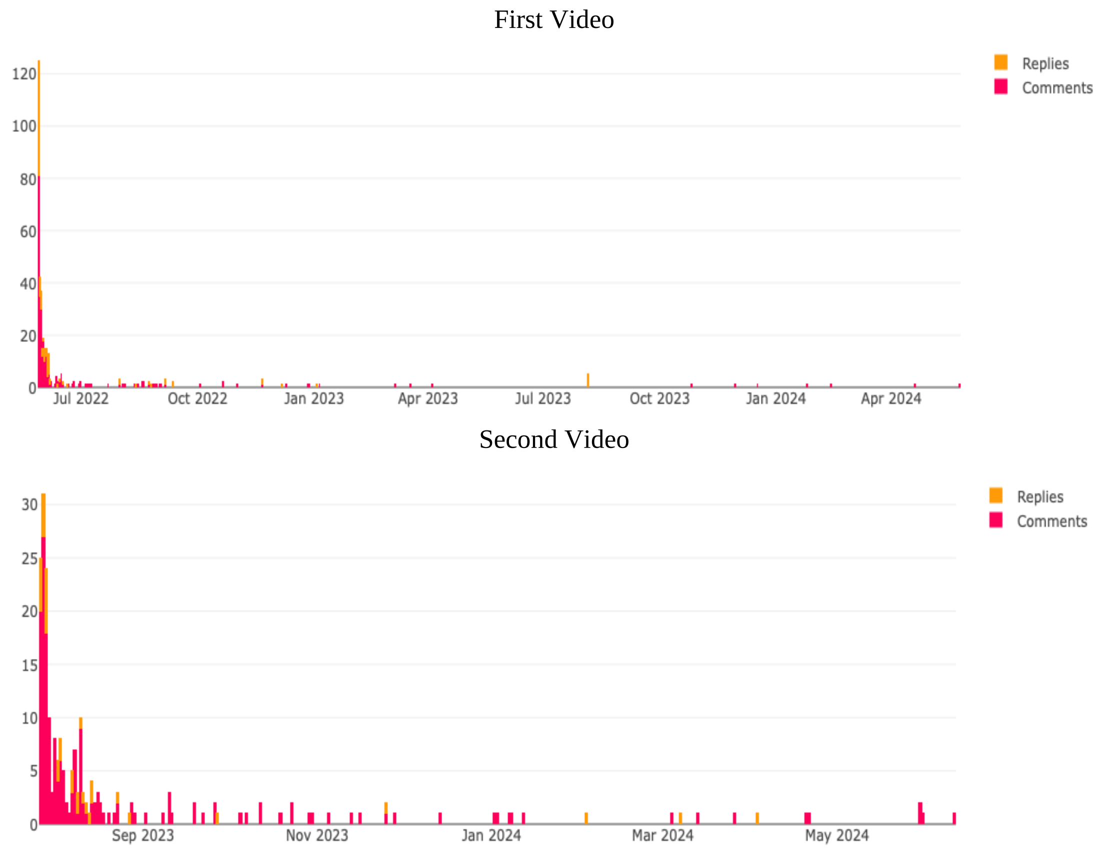


Figure 3. Video and Post-Per-Day Statistic (Communalytic)

Figure 3 shows the post-per-day statistic. Based on post-per-day statistical data, it is evident that both videos elicit comments from YouTube platform users, indicating that the travel vlog content has been viewed and well comprehended by the audience. The frequency and nature of these comments suggest active engagement and interaction, reflecting the viewers' interest and understanding of the presented material. This engagement underscores the effectiveness of travel vlogs in capturing and maintaining viewer attention. In conclusion, the observed user interactions validate the impact of travel vlog content, demonstrating its potential as an influential medium in digital tourism marketing.

The first video has accumulated 305,403 views since May 28, 2022, while the second video has garnered 253,971 views as of July 27, 2023. These substantial view counts reflect significant audience interest and engagement with the content. The high number of views indicates that the travel vlogs have effectively reached a broad audience, suggesting relevance and appeal. It is posited that such levels of engagement demonstrate the effectiveness of visual content in attracting and retaining viewer attention. In conclusion, the impressive viewership of both videos underscores the impactful role of travel vlogs in digital tourism promotion.

2.2.2 Data Understanding

In the data understanding phase, 385 data points will be processed for the video with ID s9ut2k72uwc and 220 for the video with ID S7utYxLazMA. The next crucial step is identifying the frequently used words and the top-ten posters from both videos. This analysis will reveal common themes and critical contributors, providing deeper insights into audience engagement and content reception. It is posited that understanding these elements will enhance the effectiveness of subsequent data modeling and evaluation. In conclusion, thorough data understanding lays the foundation for a robust analysis, driving more accurate and actionable insights.

Frist Video Words cloud

Second Video Words cloud

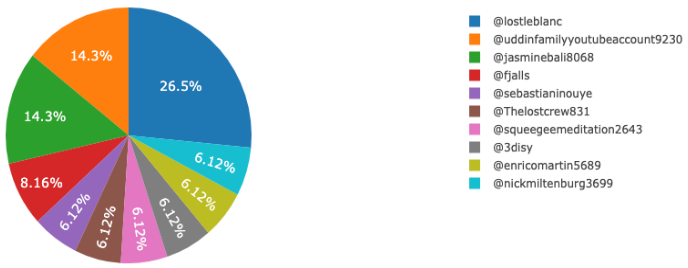
Figure 4. Words Cloud (Communalytic)

Figure 4 shows the frequently used words in the first and second videos. The most frequently occurring words in the review data of the first video are as follows: "Bali" (95 occurrences), "Gili" (87 occurrences), "Lombok" (51 occurrences), "island" (49 occurrences), "like" (36 occurrences), "video" (34 occurrences), "just" (31 occurrences), "back" (28 occurrences), and "Bali" (23 occurrences). These keywords highlight the primary locations and general sentiments expressed by viewers, indicating a strong interest in specific destinations and the content of the travel vlog. The repetition of terms such as "Bali" and "Gili" underscores the prominence of the discussions, reflecting the viewers' focus on these locations. In conclusion, analyzing these frequently used words provides valuable insights into the audience's interests and the video's impact, guiding more targeted content creation and marketing strategies.

The most frequently occurring words in the review data of the second video are as follows: "Gili" (50 occurrences), "Bali" (33 occurrences), "Lombok" (29 occurrences), "video" (24 occurrences), "back" (23 occurrences), "island" (23 occurrences), "time" (22 occurrences), "Indonesia" (21 occurrences), and "just" (18 occurrences). These words highlight key destinations and sentiments expressed by viewers, indicating a strong focus on specific locations and the overall content of the travel vlog. The prominence of terms such as "Gili" and "Bali" suggests that these destinations are central to the viewers' interest and discussions. In conclusion, analyzing these frequently used words offers valuable insights into audience engagement and the video's impact, aiding in developing more targeted and effective content strategies.

Subsequently, it is necessary to identify the top ten posters for each video to determine the viewer accounts actively commenting on the first and second videos. This analysis will reveal the most engaged audience members and the frequency of interaction, providing insights into the community dynamics surrounding the travel vlog content. Understanding these active users' engagement patterns helps tailor content to meet audience preferences better and enhance viewer retention. It is posited that recognizing and leveraging the contributions of these critical commenters will significantly boost the effectiveness of digital engagement strategies. In conclusion, identifying the top posters is crucial for optimizing content interaction and fostering a more engaged viewer base.

Frist Video



Second Video

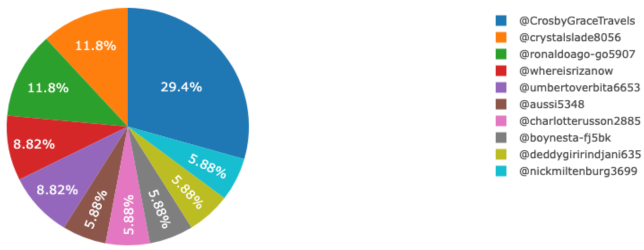


Figure 5. Top-Ten-Poster (Communalytic)

Figure 5 shows the top ten posters of the dataset. The top ten posters for the first video are identified as follows: @lostleblanc (13 comments), @uddinfamilyyoutubeaccount9230 (7 comments), @jasminebali8068 (7 comments), @fjalls (4 comments), @sebastianinouye (3 comments), @Thelostcrew831 (3 comments), @squeegeemeditation2643 (3 comments), @3disy (3 comments), @enricomartin5689 (3 comments), and @nickmiltenburg3699 (3 comments). These users demonstrate a high level of engagement, contributing significantly to the discussion around the video. The frequent interactions suggest they are particularly interested or invested in the content, highlighting critical contributors to the viewer community. It is posited that understanding the motivations and preferences of these active commenters informs more targeted and effective content strategies. In conclusion, identifying these top posters is crucial for enhancing engagement and fostering a loyal viewer base.

The top ten posters for the second video are identified as follows: @CrosbyGraceTravels (10 comments), @crystalslade8056 (4 comments), @ronaldoago-go5907 (4 comments), @whereisrizanow (3 comments), @umbertoverbita6653 (3 comments), @aussi5348 (2 comments), @charlotterusson2885 (2 comments), @boynesta-fj5bk (2 comments), @deddygiririndjani635 (2 comments), and @nickmiltenburg3699 (2 comments). These users exhibit a notable degree of engagement, actively participating in the conversation around the video. The consistent interaction signifies a strong interest in the content, marking them as key contributors within the viewer community. It is posited that analyzing the preferences and behaviors of these top commenters provides valuable insights for refining content strategies. In conclusion, recognizing these active posters is essential for enhancing audience engagement and building a dedicated viewership.

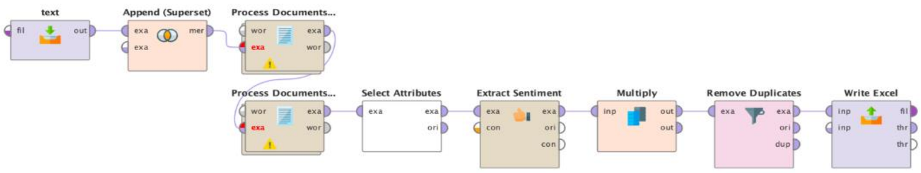


Figure 6. Data Cleaning Process (Rapidminer)

Figure 6 shows the data cleaning process in Rapidminer. Subsequently, the review data from the first and second videos will be extracted and cleansed using RapidMiner. This process systematically removes noise and irrelevant information to ensure data accuracy and reliability. Data cleansing is crucial for eliminating inconsistencies and enhancing the dataset's quality, facilitating more accurate analysis. It is posited that utilizing RapidMiner for these tasks ensures high precision and efficiency in data processing. In conclusion, the extraction and cleansing of review data are essential steps in preparing the dataset for comprehensive analysis, contributing to the validity and robustness of the research findings.

2.2.3 Modeling

the overall validity and applicability of the research findings. The cleansed and combined data will be tested using k-NN, DT, SVM, and NBC algorithms to obtain a relevant and context-appropriate model for the dataset. Each algorithm offers distinct advantages: k-NN for its simplicity and effectiveness in pattern recognition, DT for its interpretability and decision-making process, SVM for its robustness in handling high-dimensional spaces, and NBC for its efficiency in probabilistic classification. Applying these algorithms enables a comprehensive evaluation, ensuring the selection of the most suitable model. It is posited that this multi-algorithm approach enhances the accuracy and applicability of the findings. In conclusion, testing the data with these algorithms is pivotal for deriving an optimal model that aligns with the dataset's context.

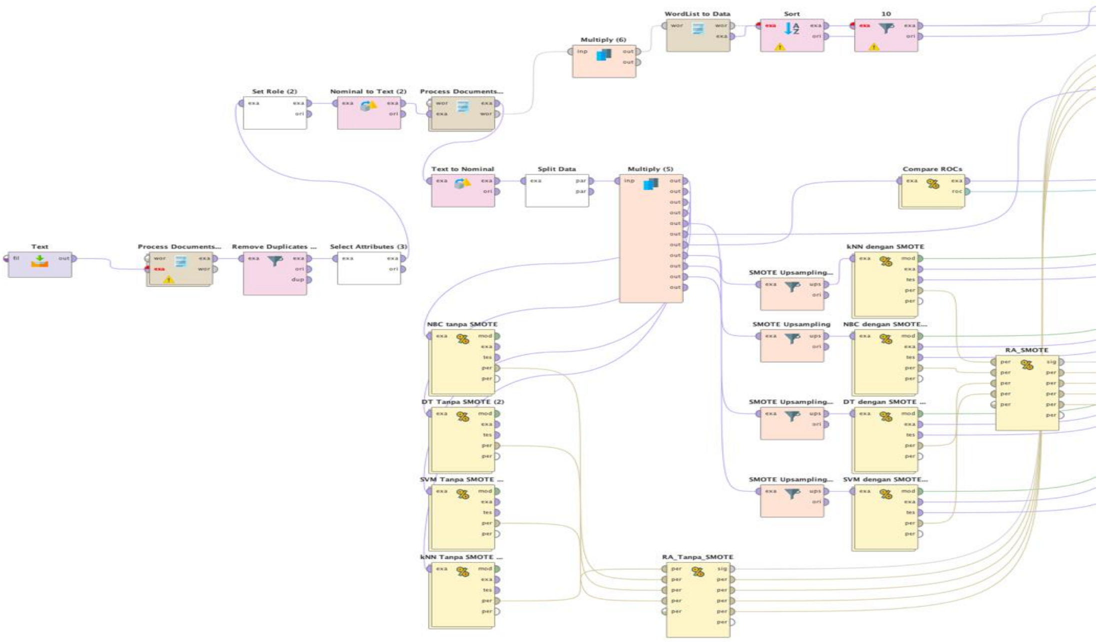


Figure 7. Implementation of SVM and DT Models in Rapidminer

Figure 7 shows the modeling process using Rapidminer. The best-performing algorithm will be evaluated based on accuracy, precision, recall, AUC, and F-measure values. These metrics provide a comprehensive assessment of the algorithm's performance: accuracy reflects the overall correctness, precision indicates the relevance of the results, recall measures the ability to identify all relevant instances, AUC assesses the algorithm's discriminatory ability, and F-measure balances precision and recall. It is posited that this multi-faceted evaluation ensures a thorough understanding of the algorithm's strengths and weaknesses. In conclusion, selecting the optimal algorithm through these metrics will ensure the reliability and effectiveness of the model in the given context.

The SMOTE operator is also employed to enhance algorithm performance, achieving suitable outputs. SMOTE, or Synthetic Minority Over-sampling Technique, addresses class imbalance by generating synthetic samples for the minority class, thus improving the algorithm's ability to learn from imbalanced data. This technique ensures that the model does not favor the majority class, leading to more balanced and accurate predictions. It is posited that using SMOTE significantly contributes to the robustness of the final model. In conclusion, incorporating the SMOTE operator is essential for optimizing algorithm performance and ensuring reliable results.

2.2.4 Evaluation

Evaluation is conducted based on the testing results of the applied models and within the context of digital engagement through travel vlog videos. The assessment involves a thorough analysis of performance metrics, ensuring the models' relevance and effectiveness in capturing viewer interaction and engagement. The study aims to derive actionable insights to enhance digital marketing strategies by aligning the evaluation with the specific context of travel vlogs. This contextual approach is posited to provide a more nuanced understanding of digital engagement dynamics. In conclusion, the evaluation process integrates model performance and contextual relevance, offering comprehensive insights into the efficacy of travel vlog content in engaging audiences.

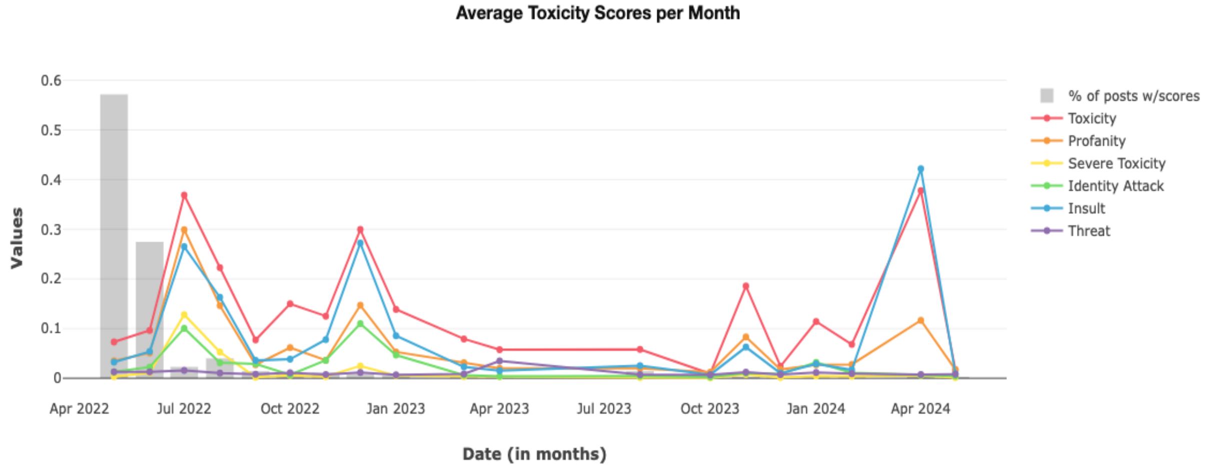
2.2.5 Deployment

Deployment represents the final phase of CRISP-DM, recommending practical measures to enhance content creator performance, digital content statistics, and viewer engagement. This stage involves implementing the insights and models developed during the previous phases, ensuring they are effectively applied to real-world scenarios. The deployment phase bridges the gap between analysis and practical application by providing actionable recommendations, enabling content creators to optimize the strategies and achieve better results. It is posited that effective deployment significantly improves the quality and impact of digital content. In conclusion, the deployment phase is crucial for translating analytical findings into tangible content creation and improving audience interaction.

3. RESULT AND DISCUSSION

Viewer responses to digital travel vlog content were categorized as positive, as indicated by the toxicity score obtained using the Perspective API from the review data on the travel vlog videos. The low toxicity scores suggest a predominance of constructive and appreciative feedback, reflecting a favorable reception among viewers. This positive engagement highlights the effectiveness of the content in resonating well with the audience and fostering a supportive community. It is posited that maintaining such positive viewer interactions is crucial for the continued success and growth of travel vlog channels. In conclusion, the toxicity score analysis confirms the positive reception of the travel vlog content, underscoring its impact and appeal.

Frist Video



Second Video

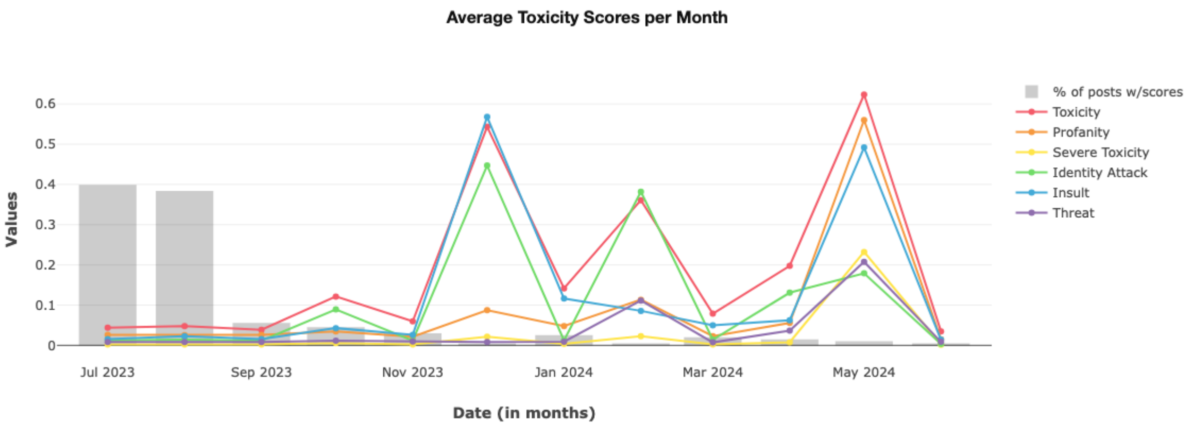


Figure 8. Average Toxicity Scores per Month

Figure 8 shows the monthly average Toxicity Scores of the first and second videos. The calculation results for the toxicity score of the first video are as follows: Toxicity 0.09627, Severe Toxicity 0.01058, Identity Attack 0.01944, Insult 0.05249, Profanity 0.05087, and Threat 0.01255. These values indicate relatively low levels of negative interaction, with toxicity and severe toxicity scores remaining below critical thresholds. The low incidence of identity attacks, insults, profanity, and threats further supports the assessment that the viewer responses are predominantly positive. It is posited that such favorable scores reflect the effectiveness of the content in maintaining a respectful and engaging viewer community. In conclusion, the low toxicity scores highlight the positive reception and constructive engagement generated by the travel vlog content.

The toxicity score calculations for the second video are as follows: Toxicity 0.06464, Severe Toxicity 0.00515, Identity Attack 0.02415, Insult 0.03163, Profanity 0.03301, and Threat 0.01144. These figures demonstrate relatively low levels of negative interaction, with the toxicity and severe toxicity scores indicating a generally positive viewer response. The minimal presence of identity attacks, insults, profanity, and threats further supports the notion that the engagement is mainly constructive. It is posited that these favorable scores reflect the content's success in fostering a respectful and engaging viewer community. In conclusion, the low toxicity scores underscore the positive reception and audience interaction the travel vlog content elicits.

Subsequently, implementing the VADER model in sentiment analysis via Communalytic for both the first and second videos indicates a predominance of positive sentiment. This analysis highlights viewers' overall favorable reception of the content, as reflected in the frequent expression of positive emotions and approval in the comments. The VADER model's ability to accurately capture and quantify these sentiments underscores the effectiveness of the travel vlog in engaging its audience. It is posited that such positive sentiment is crucial for building a supportive and loyal viewer base. In conclusion, the sentiment analysis results demonstrate the successful impact of the travel vlog content in eliciting positive viewer engagement.

Frist Video

	# of Posts	Negative Sentiment [-1..-0.05]	Neutral Sentiment (-0.05..0.05)	Positive Sentiment [0.05..1]
VADER (English/EN)	320	53 (16.56%)	83 (25.94%)	184 (57.50%)
TextBlob (English/EN)	320	38 (11.88%)	109 (34.06%)	173 (54.06%)
TextBlob (French/FR)	5	0 (0.00%)	5 (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)	170	21 (12.35%)	17 (10.00%)	132 (77.65%)
TextBlob (English/EN)	170	16 (9.41%)	38 (22.35%)	116 (68.24%)

Figure 9. Frequently used Words (Rapidminer)

Figure 9 shows the sentiment analysis using the VADER model. In the first video, based on the analysis of 329 out of 385 posts, it was found that VADER and TextBlob agreed on how to categorize 230 (71.88%) out of 320 English language posts. This level of agreement is considered moderate, as indicated by Cohen's kappa statistic of 0.516. The moderate agreement between the two sentiment analysis tools suggests a reasonable level of consistency in the sentiment categorization. It is posited that this consistency strengthens the reliability of the sentiment analysis results. In conclusion, the alignment between VADER and TextBlob provides a robust foundation for interpreting the sentiment expressed in the viewer's comments on the first video.

In the second video, based on the analysis of 170 out of 220 posts, it was found that VADER and TextBlob agree on how to categorize 127 (74.71%) out of 170 English language posts. This level of agreement is considered moderate, as indicated by Cohen's kappa statistic of 0.420. The moderate agreement between the two sentiment analysis tools indicates a reasonable level of consistency in the sentiment categorization. It is posited that this consistency supports the reliability of the sentiment analysis results. In conclusion, the alignment between VADER and TextBlob provides a solid basis for interpreting the sentiment expressed in the viewer's comments on the second video.

Specifically, in the first video, there are 27 posts (11.74%) with negative sentiments (polarity scores <= -0.05), 59 posts (25.65%) with neutral sentiments (polarity scores between -0.05 and 0.05), and 144 posts (62.61%) with positive sentiments (polarity scores >= 0.05). These figures highlight the overall positive reception of the content, with most posts expressing positive sentiments. Although less prevalent, neutral and negative sentiments provide a balanced perspective on viewer reactions. It is posited that the dominance of positive sentiments underscores the effectiveness and appeal of the travel vlog content. In conclusion, the sentiment distribution reveals a generally favorable viewer response, affirming the content's impact and resonance with its audience.

Also, in the second video, there are nine posts (7.09%) with negative sentiments (polarity scores <= -0.05), 11 posts (8.66%) with neutral sentiments (polarity scores between -0.05 and 0.05), and 107 posts (84.25%) with positive sentiments (polarity scores >= 0.05). These figures indicate an overwhelmingly positive reception of the content, with most posts expressing positive sentiments. The relatively lower presence of neutral and negative sentiments suggests that the content effectively resonates with its audience. It is posited that such a high proportion of positive feedback underscores the travel vlog's success in engaging and pleasing its viewers. In conclusion, the sentiment analysis of the second video demonstrates a highly favorable viewer response, confirming the content's strong impact.

Based on the combined data from the first and second videos, 442 data points have been cleaned, extracted, and labeled as negative or positive using RapidMiner. This comprehensive data preparation process ensures that each data point is accurately categorized, facilitating precise sentiment analysis. RapidMiner enhances the labeling process's efficiency and reliability, enabling a clear distinction between positive and negative sentiments. It is posited that such meticulous data handling is crucial for deriving valid and actionable insights. In conclusion, carefully labeling the combined dataset forms a robust foundation for subsequent sentiment analysis, contributing to the study's overall rigor and validity.

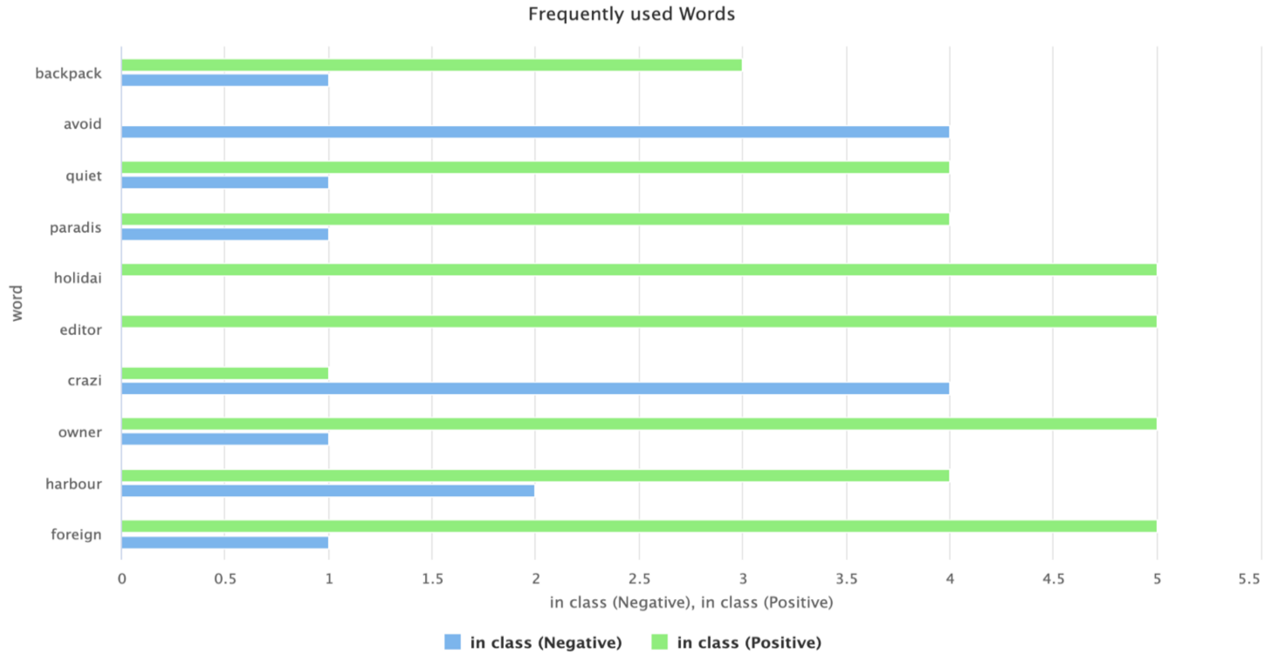


Figure 10. Frequently used Words (Rapidminer)

Figure 10 shows the frequently used words in the dataset. Based on the identification of frequently used words after the data combination, the following terms were prominent: "foreign" (4 occurrences), "harbor" (4 occurrences), "owner" (4 occurrences), "crazy" (4 occurrences), "editor" (4 occurrences), "holiday" (4 occurrences), "paradise" (4 occurrences), "quiet" (4 occurrences), "avoid" (4 occurrences), and "backpack" (2 occurrences). These keywords highlight various aspects and themes discussed in the comments, ranging from descriptions of the travel experience to specific elements of the destinations. The repeated mention of terms like "paradise" and "holiday" suggests a positive perception of the destination, while words like "avoid" and "quiet" might indicate considerations or advice for future travelers. In conclusion, analyzing these frequently used words provides valuable insights into viewer sentiments and focal points, enhancing the understanding of audience engagement with travel vlog content.

Based on the identification of frequently used words after combining the data, several key terms emerge that provide insights into viewer sentiments and focal points in the comments. The term "foreign" (4 occurrences) suggests that the content or destination appeals to or involves foreign elements, indicating significant international interest in the travel vlog. The mention of "harbor" (4 occurrences) points to specific locations or features significant to the viewers, potentially indicating popular spots or activities related to harbors in the destination. Frequent references to "owner" (4 occurrences) could imply discussions about property owners, business owners, or hosts, highlighting aspects of service or hospitality that viewers find noteworthy. The term "crazy" (4 occurrences), likely standing for "crazy," indicates exciting or unusual experiences shared by viewers, reflecting the adventurous nature of the destination.

Additionally, terms like "editor" (4 occurrences) and "holiday" (4 occurrences) suggest behind-the-scenes aspects of content creation and the vacation experience, respectively. Words like "paradise" (4 occurrences) and "quiet" (4 occurrences) indicate positive perceptions of the destination, with "paradise" suggesting idyllic scenery and "quiet" highlighting tranquility. The term "avoid" (4 occurrences) may indicate advice or warnings from viewers, while "backpack" (2 occurrences) points to a focus on travel style or target audience. In conclusion, analyzing these frequently used words provides valuable insights into viewer sentiments and focal points, enhancing the understanding of audience engagement with travel vlog content.

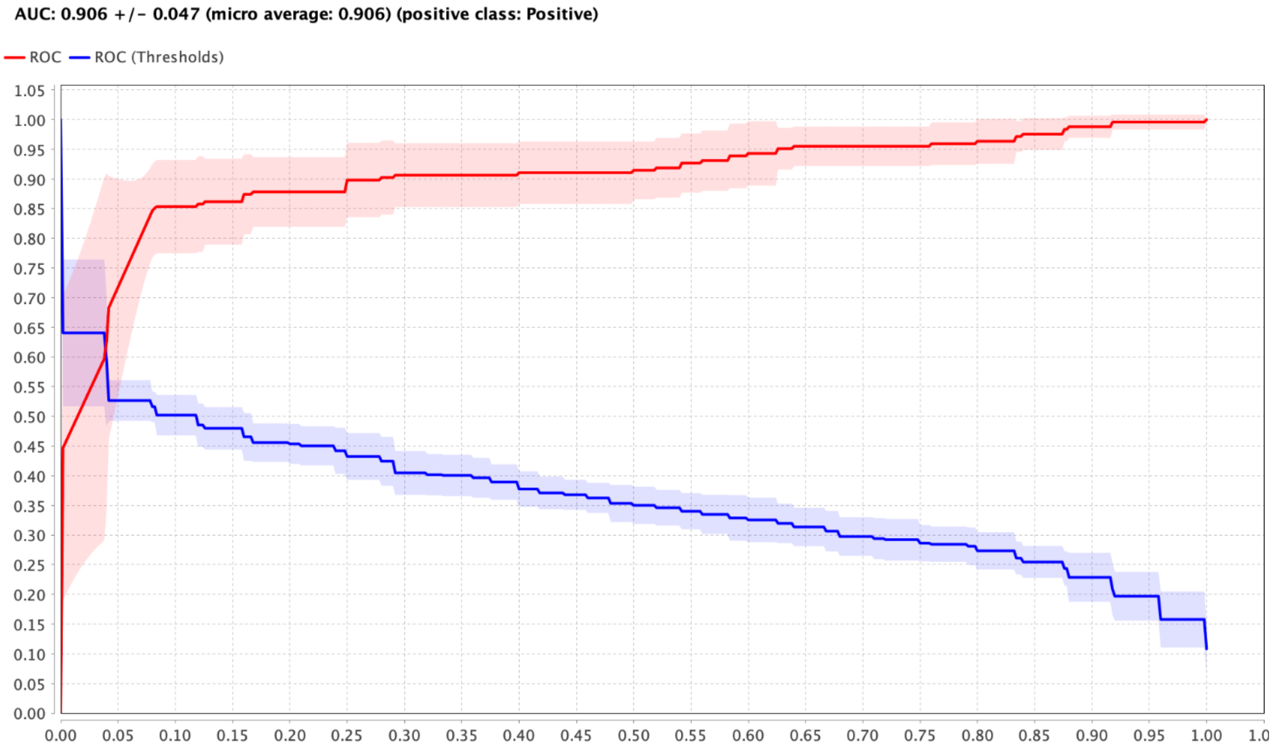


Figure 11. Area Under Curve of SVM with SMOTE in Rapidminer

Figure 11 shows the best performance of the SVM algorithm using SMOTE in the modeling phase. Subsequently, the modeling results indicate that the SVM algorithm with SMOTE demonstrated the best performance, described as follows: PerformanceVector accuracy is 88.57% +/- 5.11% (micro average: 88.57%), with the ConfusionMatrix showing 224 true negatives, 35 false negatives, 21 false positives, and 210 true positives. The AUC (optimistic) is 0.913 +/- 0.046 (micro average: 0.913), the AUC is 0.906 +/- 0.047 (micro average: 0.906), and the AUC (pessimistic) is 0.899 +/- 0.049 (micro average: 0.899), all for the positive class. The precision stands at 90.95% +/- 5.09% (micro average: 90.91%), with a recall of 85.75% +/- 6.36% (micro average: 85.71%). The f_measure is 88.22% +/- 5.30% (micro average: 88.24%), indicating a robust balance between precision and recall. It is posited that the high accuracy and precision, combined with the balanced recall and f_measure, affirm the efficacy of SVM with SMOTE in handling class imbalance and providing reliable predictions. In conclusion, the superior performance metrics of SVM with SMOTE underscore its suitability for the given dataset, ensuring accurate and consistent classification outcomes.

Thus, the VADER sentiment analysis model and the SVM algorithm with SMOTE demonstrate excellent performance in the sentiment classification process. The VADER model effectively captures the nuances of sentiment expressed in the text, providing a detailed sentiment polarity analysis. Meanwhile, the SVM algorithm with SMOTE addresses class imbalance and enhances the classification outcomes' accuracy, precision, and recall. These tools complement each other, ensuring a robust and reliable sentiment analysis framework. It is posited that the combined use significantly improves the accuracy and depth of sentiment classification. In conclusion, integrating VADER and SVM with SMOTE offers a powerful approach to achieving high-quality sentiment analysis results.

The limitations and shortcomings of this study are noteworthy. Firstly, the dataset is confined to specific travel vlogs about Gili Trawangan, which may not represent broader trends in digital tourism content. Secondly, relying on YouTube comments as the sole data source introduces potential biases related to user demographics and engagement levels. Additionally, while robust, the sentiment analysis models may not fully capture the subtleties of sarcasm, humor, or cultural nuances in the comments. It is posited that these factors could affect the generalizability and accuracy of the findings. In conclusion, while the study offers valuable insights, addressing these limitations in future research could enhance the comprehensiveness and applicability of the results.

Recommendations for the development of this research and for enhancing digital engagement through travel vlog content in tourism marketing are proposed. Future research should expand the dataset to include various destinations and diverse content platforms to ensure broader applicability and reduce potential biases. Incorporating advanced sentiment analysis techniques that better capture linguistic nuances could also improve the accuracy of the findings. For tourism marketing, leveraging the engaging nature of travel vlogs by incorporating interactive elements and real-time viewer engagement strategies is crucial. It is posited that personalized content and targeted marketing efforts, informed by detailed sentiment analysis, significantly enhance viewer retention and brand loyalty. In conclusion, these recommendations aim to advance academic research and practical applications in tourism marketing, ensuring more comprehensive and effective strategies.

4. CONCLUSION

The research concludes that applying the CRISP-DM framework in analyzing digital engagement through travel vlog content has proven effective and insightful. By systematically following the CRISP-DM phases—business understanding, data understanding, data preparation, modeling, evaluation, and deployment—the study achieved a robust analysis of viewer sentiment towards travel vlogs focused on Gili Trawangan. Specifically, the VADER sentiment analysis model and the SVM algorithm with SMOTE demonstrated excellent performance, with SVM achieving an accuracy of 88.57% +/- 5.11% and a precision of 90.95% +/- 5.09%. The analysis of 442 cleaned and labeled data points revealed that positive sentiments overwhelmingly dominated the viewer responses, accounting for 62.61% in the first video and 84.25% in the second video. These findings highlight the significant potential of travel vlogs as a powerful tool for tourism marketing, engaging a broad audience and generating positive viewer interactions. In conclusion, the CRISP-DM framework facilitated a comprehensive and detailed exploration of digital engagement, underscoring its utility in sentiment analysis and the strategic enhancement of tourism marketing efforts.

ACKNOWLEDGEMENTS

I sincerely thank the Tourism Department, the Faculty of Business Administration and Communication, LPPM, the Center of Digital Transformation and Tourism Development, and the Atma Jaya Catholic University of Indonesia.

REFERENCES

[1] S. Stensland, M. Mehmetoglu, Å. S. Liberg, and Ø. Aas, "Angling destination loyalty–A structural model approach of freshwater anglers in Trysil, Norway," *Scand. J. Hosp. Tour.*, vol. 21, no. 4, pp. 407–421, 2021, doi: 10.1080/15022250.2021.1921022.

[2] X. Guo and J. A. Pesonen, "The role of online travel reviews in evolving tourists' perceived destination image," *Scand. J. Hosp. Tour.*, vol. 22, no. 4–5, pp. 372–392, 2022, doi: 10.1080/15022250.2022.2112414.

[3] S. Lee, B. Sung, and J. Kitin, "The watchful eye: investigating tourist perceptions of different wireless tracking technologies at a travel destination," *Curr. Issues Tour.*, vol. 27, no. 6, pp. 906–922, 2024, doi: 10.1080/13683500.2023.2188582.

[4] O. A. George and C. M. Q. Ramos, "Sentiment analysis applied to tourism: exploring tourist-generated content in the case of a wellness tourism destination," *Int. J. Spa Wellness*, no. May, pp. 1–23, 2024, doi: 10.1080/24721735.2024.2352979.

[5] 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.

[6] 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.

[7] H. F. Gholipour, R. Tajaddini, and B. Foroughi, "International tourists’ spending on traveling inside a destination: does local happiness matter?," *Curr. Issues Tour.*, vol. 26, no. 12, pp. 2027–2043, 2023, doi: 10.1080/13683500.2022.2077178.

[8] B. Ericsson, M. Lerfald, and K. Overvåg, "Second homes: from family project to tourist destination–planning and policy issues," *Scand. J. Hosp. Tour.*, pp. 1–22, 2024, doi: 10.1080/15022250.2024.2332309.

[9] E. Happ, "Tourism destination competitiveness with a particular focus on sport: the current state and a glance into the future–a systematic literature analysis," *J. Sport Tour.*, vol. 25, no. 1, pp. 66–82, 2021, doi: 10.1080/14775085.2021.1888775.

[10] H. Xia, B. Muskat, M. Karl, G. Li, and R. Law, "Destination competitiveness research over the past three decades: a computational literature review using topic modelling," *J. Travel Tour. Mark.*, vol. 41, no. 5, pp. 726–742, 2024, doi:

10.1080/10548408.2024.2332278.
[11] M. Nematpour, M. Khodadadi, S. Makian, and M. Ghaffari, "Developing a Competitive and Sustainable Model for the Future of a Destination: Iran's Tourism Competitiveness," *Int. J. Hosp. Tour. Adm.*, vol. 25, no. 1, pp. 92–124, 2024, doi: 10.1080/15256480.2022.2081279.
[12] B. King, G. Richards, and E. Yeung, "City neighbourhood branding and new urban tourism," *Curr. Issues Tour.*, vol. 27, no. 10, pp. 1649–1665, 2024, doi: 10.1080/13683500.2023.2214719.
[13] R. M. Shaw and A. Ker, "Narratives of relationality and time in fertility preservation vlogs," *Fem. Media Stud.*, vol. 24, no. 3, pp. 514–530, 2023, doi: 10.1080/14680777.2023.2200591.
[14] M. Boyden, "From effluvia to chemicals: techniques of self and somatic ethics in tropical health travel narratives," *Stud. Neophilol.*, vol. 93, no. 2, pp. 155–172, 2021, doi: 10.1080/00393274.2021.1916991.
[15] C. S. Ritter, "Gazing from the air: tourist encounters in the age of travel drones," *Tour. Geogr.*, vol. 0, no. 0, pp. 1–17, 2023, doi: 10.1080/14616688.2023.2264823.
[16] N. Basaraba, "The rise of paranormal investigations as virtual dark tourism on YouTube," *J. Herit. Tour.*, vol. 19, no. 2, pp. 287–309, 2024, doi: 10.1080/1743873X.2023.2268746.
[17] E. Rosamond, "YouTube personalities as infrastructure: assets, attention choreographies and cohortification processes," *Distinktion*, vol. 24, no. 2, pp. 254–282, 2023, doi: 10.1080/1600910X.2023.2185873.
[18] M. Jaakkola, "Digital performance at the side stage: the communicative practices of classical musicians and music hobbyists on Instagram," *Continuum (N. Y).*, vol. 37, no. 2, pp. 296–308, 2023, doi: 10.1080/10304312.2023.2234109.
[19] J. Kokkranikal and E. Carabelli, "Gastronomy tourism experiences: the cooking classes of Cinque Terre," *Tour. Recreat. Res.*, vol. 49, no. 1, pp. 161–172, 2024, doi: 10.1080/02508281.2021.1975213.
[20] C. Ryan, "Mud, sweat and cameras–Irish trail and mountain running vlogging," *Sport Soc.*, vol. 27, no. 6, pp. 965–982, 2024, doi: 10.1080/17430437.2024.2334600.
[21] T. Rasul *et al.*, "Immersive virtual reality experiences: boosting potential visitor engagement and attractiveness of natural world heritage sites," *Asia Pacific J. Tour. Res.*, vol. 29, no. 5, pp. 515–526, 2024, doi: 10.1080/10941665.2024.2332993.
[22] E. Chao, "Determinants of loyalty in destination branding: perspective from supply side," *Cogent Bus. Manag.*, vol. 11, no. 1, p., 2024, doi: 10.1080/23311975.2024.2322777.
[23] C. Doppelhofer, "Overcoming the Troubles in Westeros: changing perceptions of post-conflict Northern Ireland through the diegetic heritage of Game of Thrones," *Soc. Cult. Geogr.*, vol. 25, no. 5, pp. 754–774, 2024, doi: 10.1080/14649365.2023.2209062.
[24] J. Baalbaki, "Egyptian crises and destination brand image: the resurrection of the mummy," *Curr. Issues Tour.*, vol. 27, no. 6, pp. 887–905, 2024, doi: 10.1080/13683500.2023.2187280.
[25] M. Alam, M. Kuppusamy, and P. Kunasekaran, "The role of destination management in mediating the determinants of CyberjayaTourism image branding," *Cogent Bus. Manag.*, vol. 10, no. 3, 2023, doi: 10.1080/23311975.2023.2259579.
[26] D. Amani, "The role of destination brand-oriented leadership in shaping tourism destination brand ambassadorship behavior among local residents in Tanzania," *Cogent Soc. Sci.*, vol. 9, no. 2, 2023, doi: 10.1080/23311886.2023.2266634.
[27] H. Sojasi Qeidari, S. Seyfi, C. M. Hall, T. Vo-Thanh, and M. Zaman, "'You wouldn't want to go there': what drives the stigmatization of a destination?," *Tour. Recreat. Res.*, vol. 0, no. 0, pp. 1–12, 2023, doi: 10.1080/02508281.2023.2175561.
[28] M. E. Styvén, M. M. Mariani, and C. Strandberg, "This Is My Hometown! The Role of Place Attachment, Congruity, and Self-Expressiveness on Residents' Intention to Share a Place Brand Message Online," *J. Advert.*, vol. 49, no. 5, pp. 540–556, 2020, doi: 10.1080/00913367.2020.1810594.
[29] S. Gössling, R. Vogler, A. Humpe, and N. Chen, "National tourism organizations and climate change," *Tour. Geogr.*, vol. 0, no. 0, pp. 1–22, 2024, doi: 10.1080/14616688.2024.2332368.
[30] B. Nguyen Viet, H. P. Dang, and H. H. Nguyen, "Revisit intention and satisfaction: The role of destination image, perceived risk, and cultural contact," *Cogent Bus. Manag.*, vol. 7, no. 1, 2020, doi: 10.1080/23311975.2020.1796249.
[31] U. Zaman, F. Bayrakdaroglu, M. Gohar Qureshi, M. Aktan, and S. Nawaz, "Every 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.
[32] Y. Gao, Q. Zhang, X. Xu, F. Jia, and Z. Lin, "Service design for the destination tourism service ecosystem: a review and extension," *Asia Pacific J. Tour. Res.*, vol. 27, no. 3, pp. 225–245, 2022, doi: 10.1080/10941665.2022.2046119.
[33] P. Martí, C. García-Mayor, and L. Serrano-Estrada, "Taking the urban tourist activity pulse through digital footprints," *Curr. Issues Tour.*, vol. 24, no. 2, pp. 157–176, 2021, doi: 10.1080/13683500.2019.1706458.
[34] A. H. Zins and A. Abbas Adamu, "Heritage storytelling in destination marketing: cases from Malaysian states," *J. Herit. Tour.*, vol. 0, no. 0, pp. 1–13, 2023, doi: 10.1080/1743873X.2023.2232476.
[35] R. Huang and H. M. Bu, "Destination Attributes of Memorable Chinese Rural Tourism Experiences: Impact on Positive Arousal, Memory and Behavioral Intention," *Psychol. Res. Behav. Manag.*, vol. 15, pp. 3639–3661, 2022, doi: 10.2147/PRBM.S387241.
[36] D. Amani, "It's mine, that's why I speak: Destination brand psychological ownership as a mediator of the link between social media brand engagement and local residents' voice behavior," *Cogent Bus. Manag.*, vol. 9, no. 1, 2022, doi: 10.1080/23311975.2022.2031432.
[37] 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.
[38] 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.
[39] T. T. B. Bui, "Applying the extended theory of planned behavior to understand domestic tourists' behaviors in post COVID-19 era," *Cogent Soc. Sci.*, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2166450.
[40] 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.
[41] 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.

[42] L. Chang, S. Vada, B. D. Moyle, and S. Gardiner, "Unlocking the gateway to tourist well-being: the Triple-A model of engagement in tourism experiences," *Curr. Issues Tour.*, no. May, pp. 1–20, 2024, doi: 10.1080/13683500.2024.2359544.

[43] 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.

[44] 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.

[45] 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.

[46] E. Agyeiwaah and Y. Zhao, "A moderated mediation model of tourist cultural worldviews, perceived social relations, self-esteem, and prosocial behaviors - analyzing competing models," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–21, 2023, doi: 10.1080/09669582.2023.2253501.

[47] 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.

[48] S. W. Lin, K. F. Wang, and Y. H. Chiu, "Effects of tourists' psychological perceptions and travel choice behaviors on the nonmarket value of urban ecotourism during the COVID-19 pandemic- case study of the Maokong region in Taiwan," *Cogent Soc. Sci.*, vol. 8, no. 1, 2022, doi: 10.1080/23311886.2022.2095109.

[49] N. Shen, H. Zhang, H. Wang, X. Zhou, L. Zhou, and G. Tang, "Toward multi-granularity spatiotemporal simulation modeling of crowd movement for dynamic assessment of tourist carrying capacity," *GIScience Remote Sens.*, vol. 59, no. 1, pp. 1857–1881, 2022, doi: 10.1080/15481603.2022.2139450.

[50] J. Zhou, K. Xiang, Q. Cheng, and C. Yang, "Psychological and behavioural consistency value seeking of tourists in niche tourism: Nostalgia, authenticity perception, and satisfaction," *Psychol. Res. Behav. Manag.*, vol. 14, pp. 1111–1125, 2021, doi: 10.2147/PRBM.S322348.