KLIK: Kajian Ilmiah Informatika dan Komputer
ISSN 2723-3898 (Media Online)
Vol 5, No 1, August 2024, Hal 142-153
DOI 10.30865/klik.v5i1.2091
https://djournals.com/klik

Sentiment and Toxicity Analysis of Digital Content Using Perspective, Vader, and TextBlob: Tourism and Birdwatching

Yerik Afrianto Singgalen

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

Abstract–This research investigates the impact of digital content on specialized tourism activities, focusing on birdwatching, using tools such as Communalytic and RapidMiner. By analyzing 1,021 posts, the study reveals an average toxicity score of 0.13839, with VADER identifying 32.78% negative sentiment and TextBlob identifying 17.07% negative sentiment. Despite these negative interactions, over 50% of the posts convey positive sentiment, highlighting the potential for digital content to foster a supportive and engaging community. The findings underscore the urgent need to address toxicity to maintain a positive online environment, crucial for enhancing educational outreach and participant engagement. This research emphasizes the critical and immediate role of digital platforms, analyzed through Communalytic and RapidMiner, in promoting environmental awareness and conservation, thereby driving the growth and sustainability of niche tourism sectors such as birdwatching. Prompt action is essential to leverage these insights to benefit the environment and local economies.

Keywords: Sentiment; Toxicity; Digital Content; Perspective; Vader; TextBlob; Birding

1. INTRODUCTION

Birding, or birdwatching, constitutes observing avian species within their natural habitats, forming a niche tourism segment gaining global popularity. This activity appeals to a wide range of individuals, from casual hobbyists to dedicated ornithologists, drawn by the allure of witnessing birds in their authentic environments [1]–[3]. The growing interest in birdwatching reflects a broader trend toward eco-friendly and immersive travel experiences [4]–[7]. As a result, birding contributes to environmental awareness and conservation efforts and enhances the appreciation of biodiversity, underscoring its value as a significant component of specialized tourism.

Birding activities captivate a diverse audience, ranging from amateurs to experts, facilitated by digital technology that enables broader participation through social media platforms and online communities. This technological integration allows enthusiasts to share sightings, discuss species, and collaborate on identification, enriching the birdwatching experience [8]–[12]. Such digital tools significantly enhance engagement and accessibility, making birding a more inclusive and interactive hobby [13]. The intersection of birding and digital innovation broadens its appeal and fosters a global network of avian enthusiasts, highlighting the transformative impact of technology on traditional recreational activities.

This study aims to provide a comprehensive overview of how negative digital interactions influence birding activities and to suggest measures for improving the quality of the birdwatching experience in digital spaces. The examination identifies harmful interactions and their impact on participants' engagement and enjoyment. By understanding these dynamics, the research proposes targeted strategies to mitigate negative influences, thereby fostering a more positive and supportive online environment. Ultimately, this study seeks to enhance the overall experience for birdwatching enthusiasts by alleviating and easing the challenges posed by digital interactions [14], [15]. The primary objective of this study is to examine sentiment and toxicity levels within digital content related to birding, thereby identifying social dynamics and adverse impacts within online birdwatching communities. Analyzing these elements involves assessing the emotional tone of posts and comments and measuring the prevalence of harmful language. This investigation highlights the importance of fostering a positive and inclusive environment in digital spaces, as negative interactions can detract from birds' educational and recreational value. Understanding and mitigating toxicity in online content is crucial for maintaining a supportive and engaging community for birding enthusiasts.

Understanding the darker aspects of birding is crucial, as this research aims to uncover harmful elements such as verbal harassment, user conflicts, and the spread of misinformation, all of which can impact the experience and participation in this niche tourism activity. These issues often remain obscured yet significantly affect birdwatching communities' overall dynamics and enjoyment [16]–[18]. Addressing these problems is essential for fostering a safe and supportive environment for all participants. Thus, this study highlights the importance of recognizing and mitigating these negative influences to enhance the quality and inclusivity of birding as a specialized form of tourism.

This research will enrich the literature on social dynamics within birdwatching communities and enhance understanding of how negative sentiment and toxic behavior evolve in digital contexts. By analyzing these phenomena, the study will provide valuable insights into the underlying factors contributing to such adverse interactions [19]. Consequently, the findings will inform the development of more effective mitigation strategies to foster a more positive and inclusive online environment for the birding community. This research addresses the current challenges and creates a more supportive digital space for enthusiasts.

Previous studies have investigated sentiment and toxic behavior within other hobby communities, such as gaming or travel, to understand patterns and characteristics that might be similar in the context of birdwatching [20]. These

investigations have revealed commonalities in the emergence and propagation of negative behaviors driven by anonymity, competitive dynamics, and the communal nature of online platforms [21]. Recognizing these parallels allows for a more nuanced analysis of the birdwatching community, offering a basis for targeted interventions [22]. Thus, this comparative approach underscores the importance of addressing toxic behavior across various digital hobbyist environments to foster healthier interactions.

Limitations concerning data access, population representation, and potential biases in sentiment analysis reliant on natural language processing algorithms constrain this research. These constraints may affect the comprehensiveness and accuracy of the findings, as the data available might not fully capture the diversity of the birdwatching community. Moreover, the algorithms used for sentiment analysis can introduce biases, potentially skewing the interpretation of sentiment and toxicity levels. Despite these limitations, the study provides valuable insights, highlighting the need for continuous refinement of methodologies to achieve more accurate and representative results in future research.

2. RESEARCH METHODOLOGY

2.1 Gap Analysis

Identifying research gaps in previous studies on birdwatching and tourism is essential for advancing scholarly understanding and guiding future investigations. Recognizing these gaps enables researchers to pinpoint areas that have been underexplored, such as the socio-economic impacts of birdwatching tourism or the role of digital platforms in shaping birdwatching behaviors [23]. Addressing these unexplored dimensions enriches the literature and provides practical insights for stakeholders in the tourism industry. Consequently, thoroughly identifying research gaps fosters a more comprehensive and nuanced approach to studying birdwatching within the broader context of tourism, ultimately enhancing both academic and practical outcomes.

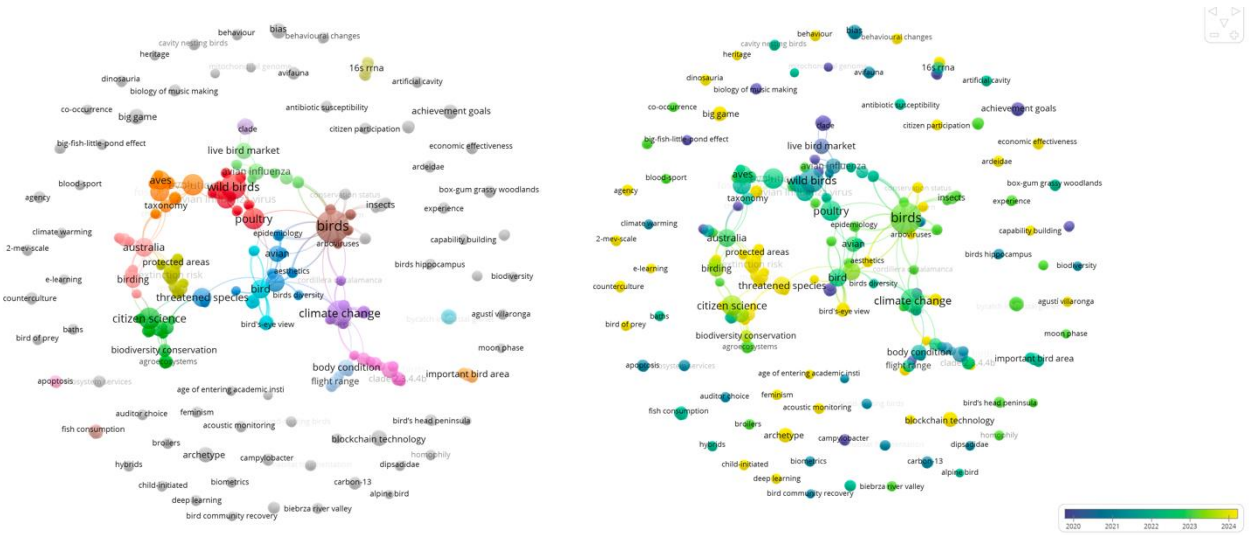

Figure 1. Network and Overlay Visualization

Figure 1 shows network and overlay visualization using Vosviewer. The interconnected nature of research topics within birding highlights the multidisciplinary approach required to understand this field comprehensively. Key areas of study such as climate change, biodiversity conservation, and citizen science interact closely with more specialized topics like avian influenza, live bird markets, and the impact of technology on birdwatching [24]. These connections suggest that advances in one area significantly influence findings in others, underscoring the importance of an integrated research strategy. Therefore, fostering collaboration across these diverse yet interconnected topics can enhance birding research's overall understanding and effectiveness, ultimately contributing to scientific knowledge and practical applications in conservation and tourism.

The novelty of this research, as evidenced by the network and overlay visualization of previous studies related to birding, lies in its integrative approach to examining biological and socio-technological dimensions. This research uniquely bridges gaps by incorporating the impacts of climate change, biodiversity conservation, and citizen science with emerging digital platforms and technologies. By doing so, it addresses underexplored intersections, such as the influence of social media on birdwatching behaviors and the role of technological advancements in enhancing bird conservation efforts. This comprehensive perspective advances theoretical understanding and offers practical solutions for promoting sustainable birdwatching practices, marking a significant contribution to the field.

2.2 Digital Content Reviews and Analysis Framework

The Digital Content Reviews and Analysis Framework can transform textual data from digital content comments into actionable insights for analyzing the context of tourism and birdwatching. This framework systematically processes raw

textual data through data cleaning, extraction, and evaluation stages, ensuring high-quality inputs for subsequent analysis. The framework identifies key themes and sentiments within the comments by employing advanced natural language processing techniques, revealing valuable patterns and trends. Consequently, this methodological approach enhances the understanding of viewer engagement and sentiment, providing a robust foundation for strategic decision-making in the tourism and birdwatching sectors.

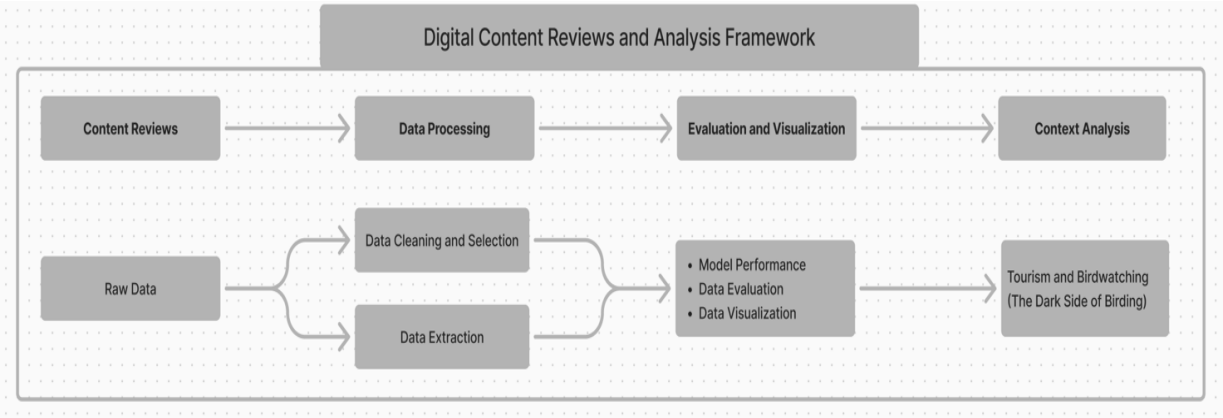

Figure 2. Digital Content Reviews and Analysis Framework

Figure 2 shows the digital content reviews and analysis framework. This research employs the Digital Content Reviews and Analysis Framework, which encompasses stages of content reviews, data processing, evaluation and visualization, and context analysis. Initially, content reviews systematically collect and assess relevant digital material, followed by data processing that involves cleaning, selection, and extraction to ensure high-quality data inputs. Subsequently, the evaluation and visualization phase interprets data patterns and model performance, providing insightful visual representations. This rigorous process culminates in context analysis, offering a nuanced understanding of the digital interactions surrounding birdwatching and tourism. Ultimately, this framework facilitates a comprehensive digital content examination, enhancing the research findings' quality and depth.

The data sources for this study are derived from a YouTube platform with the ID G_zTHrFy7c0, which has garnered 293,180 views since March 20, 2023. Additionally, 1,021 comments reflect the viewers' engagement with the video content. These metrics indicate substantial interaction and interest, providing a rich dataset for analyzing viewer responses and sentiment. The significant volume of comments offers valuable insights into the audience's perceptions and reactions, making it a pertinent source for examining digital engagement within the birdwatching community. Consequently, this data is a robust foundation for understanding the social dynamics and impact of digital content in this niche.

2.2.1 Content Reviews

The content with ID G_zTHrFy7c0 exhibits a detailed post-per-day statistic, reflecting varying levels of engagement over time. For instance, the highest activity was recorded on March 20, 2023, with 29 posts, followed by significant fluctuations in the subsequent days and months. This interaction pattern demonstrates the dynamic nature of viewer engagement and highlights periods of heightened interest and discussion. The analysis of these statistics offers valuable insights into the temporal distribution of digital interactions, suggesting potential correlations between posting frequency and viewer engagement. Such data is instrumental in understanding the ebb and flow of digital content consumption and its impact on community participation.

In the initial stage, it is essential to ascertain the post-per-day statistic content data to determine the number of comments. It replies from viewers, serving as a representation of the response to the content. This metric provides a quantitative measure of engagement, illustrating the frequency and intensity of interactions over time. Analyzing these statistics allows for a deeper understanding of viewer engagement patterns, highlighting periods of high activity and potential drivers of discussion. Consequently, this data forms a crucial foundation for evaluating digital content's overall impact and reception, enabling more targeted and effective content strategies.

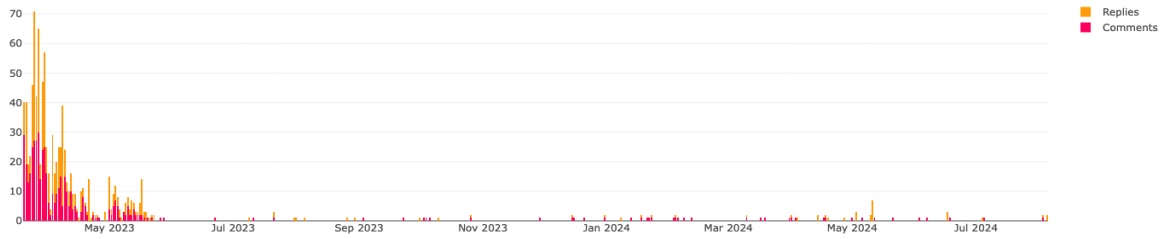

Figure 3. Post-Per-Day Statistic of the Content

Figure 3 shows the post-per-day statistic. Analyzing the post-per-day statistics for the content with ID G_zTHrFy7c0 reveals several key insights into viewer engagement and interaction trends. Initially, there was a notable spike in activity on March 20, 2023, with 29 posts indicating a high level of interest or a specific event driving engagement on that day. Following this peak, there is a significant drop and fluctuation in the number of posts per day, with subsequent notable peaks on March 27, 2023 (30 posts), March 25, 2023 (27 posts), and March 24, 2023 (25 posts). This pattern suggests periodic surges in engagement, possibly driven by new content releases, social media shares, or discussions within the community. As the data progresses into April and beyond, the number of daily posts generally declines, indicating a tapering off of initial enthusiasm or a shift in viewer interest. However, intermittent peaks, such as 34 posts on April 8, 2023, and 32 posts on March 30, 2023, suggest sporadic events or triggers that re-ignite viewer activity. The extended tail of low activity from May onwards, with occasional single-digit posts, indicates a sustained but reduced level of engagement, typical of the long-term digital content lifecycle. This long tail highlights the enduring interest among a smaller, dedicated audience segment despite the overall decline in activity. In conclusion, the data analysis indicates a highly dynamic and event-driven pattern of engagement, with initial high activity followed by periodic peaks and a long-term decline. Understanding these trends can inform content creators and community managers about optimal times for new releases, potential triggers for engagement, and strategies for sustaining viewer interest over the long term.

Subsequently, it is essential to identify the community's top ten posters to understand better the key contributors driving engagement and interaction. Recognizing these individuals provides valuable insights into the core group that significantly influences the discourse and activity levels. This identification allows for targeted analysis of their posting behavior, interaction patterns, and the content they generate. Such an analysis can reveal critical trends and factors that contribute to high levels of engagement, informing strategies to encourage broader participation. Ultimately, understanding the role and impact of these top contributors can enhance community management and foster a more dynamic and inclusive environment.

Figure 4. Top Ten Poster

Figure 4 shows the top ten posters of the content. Based on the data from the top ten posters, it is evident that @BadgerlandBirding is the most active contributor, with 75 posts significantly surpassing the other top contributors. The subsequent most active users include @pixazelz with 27 posts and @neonwhitea.1548 with 14 posts, followed by @rovingwithnature, @173jaSon371, and @audreyboag7670, who have posted 12, 11, and 9 times respectively. These users' relatively high activity level indicates a core group of engaged participants who drive much of the conversation and interaction. This concentration of posts among a few users suggests the presence of key influencers within the community, whose interactions and contributions can significantly shape the discourse and engagement levels. Understanding the dynamics of these top contributors can provide insights into community behavior and help tailor strategies to foster broader participation and enhance community engagement.

Based on the data context, it is evident that the textual data pertains to birding and requires cleaning and extraction to produce relevant information. The initial step involves removing extraneous or irrelevant content to ensure the dataset's quality and focus. The extraction process then isolates significant themes and patterns pertinent to the research objectives. By refining the data through these stages, the analysis can yield more accurate and meaningful insights, ultimately enhancing the understanding of birding-related interactions and trends within the digital content. This systematic approach is crucial for transforming raw data into valuable and actionable knowledge.

2.2.2 Data Processing: Cleaning, Selection and Extraction

The application employed for data cleaning and extraction processes is RapidMiner. This robust platform facilitates the systematic removal of irrelevant data and the extraction of pertinent information, ensuring a high-quality dataset for analysis. Utilizing RapidMiner efficiently handles large volumes of data, streamlining the pre-processing phase and enhancing the accuracy of subsequent analyses. Consequently, using RapidMiner significantly contributes to the integrity and reliability of the research findings, providing a solid foundation for insightful and actionable conclusions.

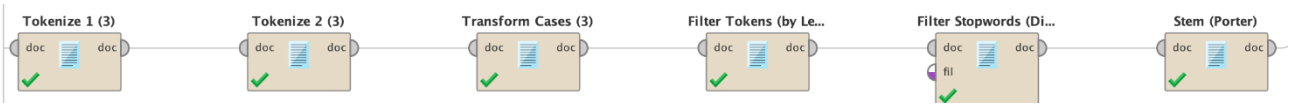

Figure 5. Cleaning Process

Figure 5 shows the cleaning process. 1,022 textual data entries were meticulously cleaned using a series of operators: tokenize, transform cases, filter tokens, filter stopwords, and stem. The tokenized operators initially divided the text into manageable units and then transformed cases to ensure uniformity. Subsequent filtering removed irrelevant tokens and common stopwords, enhancing the dataset's focus and relevance. The stemming process then reduced words to their base forms, further refining the data. This systematic cleaning ensures the dataset's integrity and accuracy, laying a solid foundation for in-depth analysis and reliable research outcomes.

The primary objective of cleaning textual data in RapidMiner is to enhance the accuracy and relevance of the dataset for subsequent analysis. This process involves several steps, including tokenization, case transformation, and removing stopwords and irrelevant tokens, collectively improving data quality. The data becomes more structured and focused by eliminating noise and inconsistencies, facilitating more precise and reliable insights. Consequently, the refined dataset allows for more practical application of analytical models, ultimately leading to more robust and meaningful research findings. This meticulous data preparation is crucial for ensuring the validity and reliability of analytical outcomes.

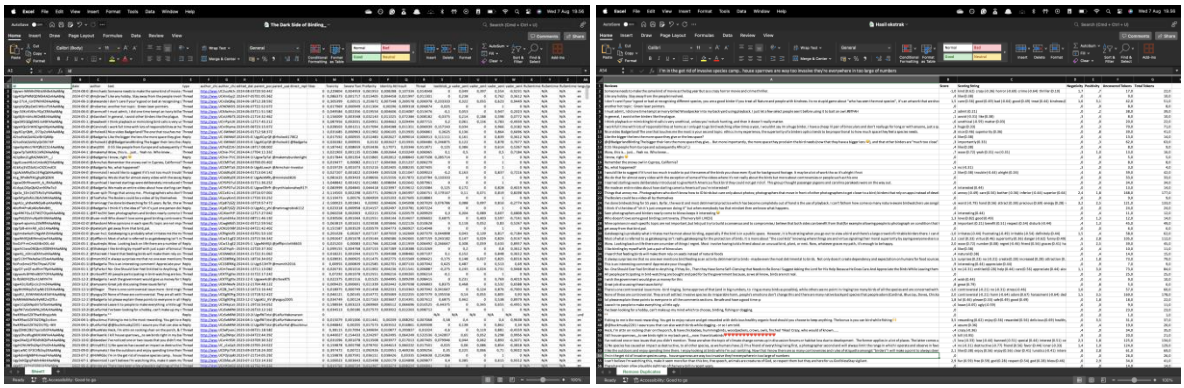

Figure 6. Data Selection Process

Figure 6 shows the data selection process. This pre-processing ensures that the dataset is devoid of noise and irrelevant information, enhancing its suitability for detailed analysis. The subsequent extraction process can be conducted with greater precision and accuracy by preparing the data through comprehensive cleaning. The availability of such a refined dataset facilitates the identification of significant patterns and insights, thereby contributing to more robust and reliable analytical outcomes. This preparatory step is crucial for achieving meaningful and actionable results in the extraction process.

In the extraction process, the data comprises 1,008 textual entries that have undergone thorough cleaning using RapidMiner. This pre-processing step involves removing irrelevant information and standardizing the text to ensure its suitability for detailed analysis. The dataset is refined by employing RapidMiner's robust data-cleaning capabilities to enhance accuracy and relevance. The cleaned data facilitates more precise extraction of meaningful patterns and insights, contributing to robust and reliable analytical outcomes. Consequently, the meticulous preparation of the dataset is fundamental to achieving significant and actionable results in the extraction process.

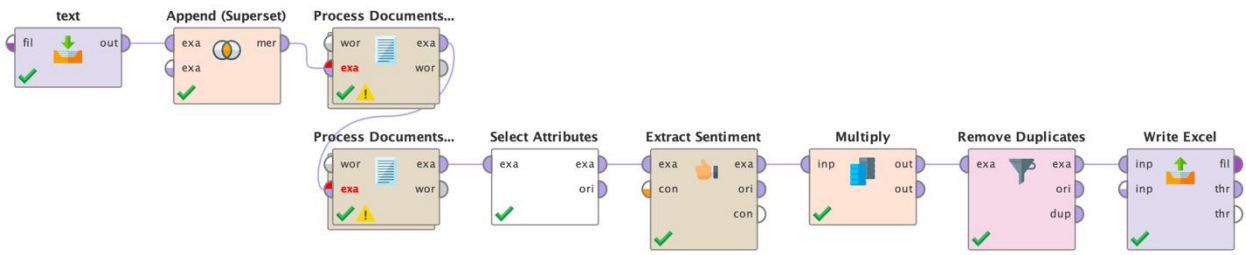

Figure 7. Extraction Process Using Vader Operator in Rapidminer

Figure 7 shows the extraction process using the Vader operator in Rapidminer. The extraction process using VADER (Valence Aware Dictionary and Sentiment Reasoner) on RapidMiner involves systematically analyzing sentiment within textual data. Initially, the text data is processed and appended to form a superset, followed by document processing to prepare the data for sentiment analysis. The VADER tool is then utilized to extract sentiment attributes, providing insights into the emotional tone of the content. This process includes selecting relevant attributes, multiplying

data sets, and removing duplicates to ensure accuracy and consistency. The processed data is ultimately exported to an Excel file, facilitating further analysis and interpretation. This comprehensive method enhances the reliability of sentiment analysis, contributing valuable information for subsequent research and decision-making.

The primary goal of extracting sentiment from textual data in RapidMiner using VADER is to identify and quantify the emotional tone of the content accurately. VADER, a robust sentiment analysis tool, leverages a lexicon-based approach to discern positive, negative, and neutral sentiments within the text. By applying this tool, the analysis can reveal underlying emotional patterns and trends, offering valuable insights into the attitudes and opinions expressed. This process enhances the understanding of audience reactions and engagement, providing a deeper contextual interpretation of the data. Ultimately, sentiment extraction aids in informed decision-making and strategy development based on the nuanced emotional landscape of the textual content.

2.2.3 Data Evaluation and Visualization

Based on the processed results of 982 posts out of 1,021 using the Perspective API, the data visualization provides a comprehensive overview of sentiment and interaction trends within the dataset. This visualization highlights the distribution and intensity of various sentiments, enabling a deeper understanding of the underlying emotional tone of the posts. By leveraging the analytical power of the Perspective API, the visualization elucidates key patterns and anomalies, offering valuable insights into user behavior and engagement. Such detailed visual representation is instrumental in identifying areas for further investigation and informing strategic decisions. Consequently, this visualization is crucial for effectively interpreting and utilizing the processed data.

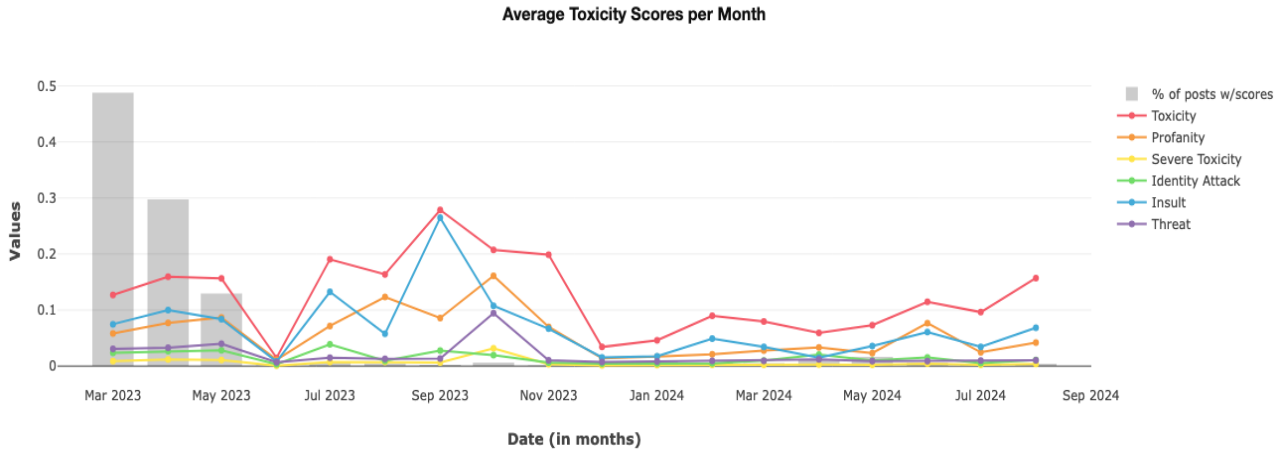

Figure 8. Visualization of Average Toxicity Score Per Month

Figure 8 shows the visualization of the average toxicity score per month. The data visualization of average monthly toxicity scores reveals significant insights into the prevalence of various negative interactions within the dataset. The average toxicity score is 0.13839, with severe toxicity averaging at 0.00979, indicating relatively low but notable occurrences of harmful content. Identity attacks and insults have average scores of 0.02433 and 0.08158, respectively, highlighting the presence of targeted negative behavior. Profanity and threats, with scores of 0.06670 and 0.03141, respectively, further illustrate the diverse forms of toxicity present. This analysis underscores the need for targeted strategies to mitigate these behaviors, enhancing the overall quality and safety of the online community. Effective intervention measures are essential to reduce these toxicity levels and foster a more positive and inclusive digital environment.

The visualization of the distribution of polarity values, based on the analysis of 968 out of 1,021 posts, reveals significant insights into the sentiment dynamics within the dataset. For VADER (English/EN), 961 posts were analyzed, with 315 (32.78%) showing negative sentiment, 165 (17.17%) exhibiting neutral sentiment, and 481 (50.05%) reflecting positive sentiment. TextBlob (English/EN) analysis of the same number of posts showed a slightly different distribution: 164 (17.07%) negative, 293 (30.49%) neutral, and 504 (52.45%) positive. In contrast, TextBlob's analysis of French (FR) and German (DE) posts showed that all French posts were equally split between neutral and positive sentiments, while all German posts were neutral. These findings indicate that most posts tend to be positive, with a significant portion exhibiting neutrality and a smaller fraction displaying negativity. This sentiment distribution underscores the overall positive engagement within the analyzed content, providing a foundation for further targeted sentiment analysis and community management strategies.

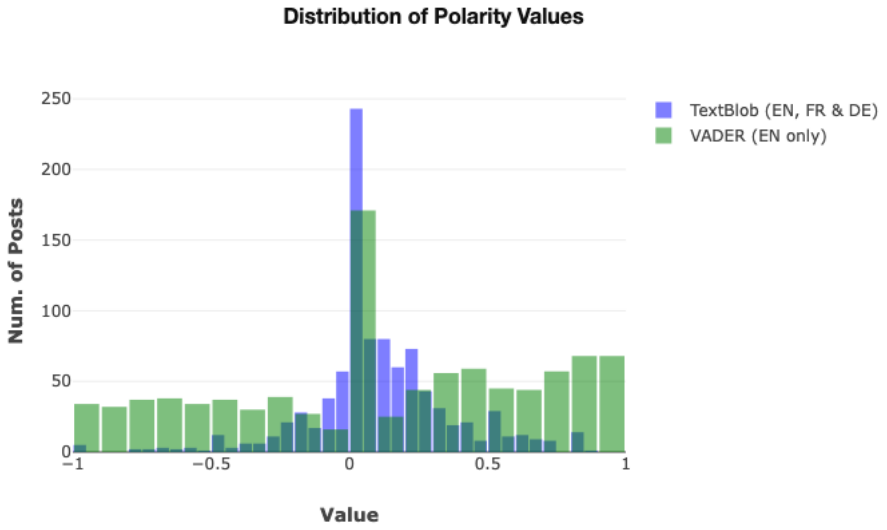

Figure 9. Distribution of Polarity Values

Figure 9 shows the distribution of polarity values. Excluding duplicates such as reposts or retweets, VADER and TextBlob concur on the categorization of 563 out of 959 English language posts, representing an agreement level of 58.71%. This level of concordance is quantified as fair, with a Cohen's kappa statistic of 0.343, indicating a moderate level of agreement between the two sentiment analysis tools. This statistic reflects the consistency in sentiment categorization between VADER and TextBlob, highlighting their relative reliability in analyzing textual data. Nonetheless, the fair agreement level suggests room for improvement in the alignment of these tools, which could be addressed through further calibration and refinement. Ultimately, achieving higher agreement levels would enhance the robustness and accuracy of sentiment analysis outcomes.

Expressly, both libraries agree on the categorization of sentiments as follows: 109 posts (19.36%) with negative sentiments, having polarity scores of -0.05 or lower; 97 posts (17.23%) with neutral sentiments, with polarity scores ranging between -0.05 and 0.05; and 357 posts (63.41%) with positive sentiments, indicated by polarity scores of 0.05 or higher. This concordance between VADER and TextBlob underscores their alignment in identifying sentiment trends within the dataset. Such agreement enhances the credibility of the sentiment analysis, suggesting a reliable classification of emotional tones across the analyzed posts. Consequently, this consistency supports the validity of the findings and informs further research and strategic decision-making.

2.2.4 Context Analysis: Tourism and Birdwatching

The identified toxicity scores and sentiment analysis data correlate with tourism and birdwatching concepts in the context analysis. This connection allows an in-depth understanding of how digital content's negative interactions and emotional tones impact the perception and engagement in these activities. The analysis reveals critical insights into community dynamics and the potential deterrents to participant involvement by linking toxicity and sentiment metrics to tourism and birdwatching. These findings highlight the necessity for targeted interventions to mitigate negative influences and enhance the overall experience for enthusiasts. Consequently, this comprehensive approach informs strategies to foster a more positive and inclusive environment in the tourism and birdwatching sectors.

Figure 10. Wordclouds of Dataset

Figure 10 shows the word clouds of the dataset. Based on the frequently used words identified in the word cloud, several key aspects need to be analyzed in the context of tourism and birdwatching. Prominent terms such as "conservation," "habitat," and "species" highlight the importance of environmental and ecological considerations in birdwatching activities. Additionally, words like "community," "experience," and "engagement" underscore the social and participatory dimensions crucial for enhancing tourist involvement. Analyzing these elements provides valuable insights into the factors driving interest and participation in birdwatching and potential challenges. Ultimately, this targeted analysis informs strategies to promote sustainable practices and enrich the overall experience for birdwatching enthusiasts within the tourism sector.

The intricate relationship between tourism and birding reveals significant environmental and economic sustainability implications. Birding as a form of ecotourism promotes the conservation of natural habitats and biodiversity by raising awareness and generating funding for preservation efforts. Economically, birding attracts a dedicated group of tourists who contribute to local economies through spending on travel, accommodation, and guided tours. This dual impact underscores the importance of integrating sustainable practices in tourism and conservation strategies. Fostering a symbiotic relationship between birding and tourism can enhance environmental stewardship while bolstering economic resilience in local communities.

3. RESULT AND DISCUSSION

The discussion in this research is divided into two main sections: toxicity score and sentiment classification of digital content. The first section delves into the analysis of toxicity scores, examining the prevalence and patterns of harmful interactions within the dataset. This analysis identifies key trends and underlying factors contributing to negative behavior in digital platforms. The second section focuses on sentiment classification, categorizing the emotional tone of the content and exploring its implications for user engagement and community dynamics. Together, these sections provide a comprehensive understanding of the digital environment, offering insights into the extent of toxicity and the overall sentiment landscape. This dual analysis is crucial for developing strategies to enhance the quality of digital interactions and foster a more positive online community.

3.1 Toxicity and Sentiment Classification of Digital Content: The Dark Side of Birding

Based on the analysis of 982 posts (out of 1,021) using the Perspective API, the results indicate varying toxicity levels within the dataset. The average toxicity score is 0.13839, with severe toxicity averaging 0.00979, highlighting the presence of some extreme negative interactions. Identity attacks have an average score of 0.02433, while insults and profanity average 0.08158 and 0.06670, respectively. Additionally, threats have an average score of 0.03141, indicating occasional hostile interactions. The highest recorded values for these categories underscore the potential severity of negative behavior in digital content. These findings emphasize the importance of addressing toxicity to foster a healthier online environment, thereby improving user experience and community engagement.

Interpreting the meaning of the toxicity score involves understanding its implications for online interactions and community health. The toxicity score quantifies the extent of harmful language within digital content, with higher scores indicating more severe negative behavior. For example, an average toxicity score of 0.13839 suggests that while most interactions are relatively benign, there are notable harmful language that could affect user experience. High values in specific categories, such as severe toxicity and identity attacks, highlight the presence of particularly damaging interactions. These metrics are crucial for identifying areas where intervention is needed to reduce toxicity and promote a more positive and inclusive online environment. Addressing high toxicity scores is essential for maintaining the integrity and well-being of digital communities.

	Average for dataset	Highest value
Toxicity	0.13839	0.83647
Severe Toxicity	0.00979	0.35292
Identity Attack	0.02433	0.54997
Insult	0.08158	0.82043
Profanity	0.06670	0.86331
Threat	0.03141	0.61642

Figure 11. Toxicity Score Result (Based on Perspective)

Figure 11 shows the toxicity score based on the Perspective Model. Connecting the interpretation of toxicity scores with digital content engagement and issues in tourism and birdwatching reveals critical insights into community dynamics. High toxicity scores can deter user participation, diminishing overall engagement and negatively impacting

the sense of community. In tourism and birdwatching, such negativity can discourage potential enthusiasts from joining online discussions, limiting the sharing of valuable information and experiences. Addressing these toxicity issues is paramount to fostering a welcoming environment that encouraging active participation and collaboration. Enhancing the quality of digital interactions will improve user satisfaction and support the growth and sustainability of tourism and birdwatching communities. Therefore, mitigating toxicity is essential for nurturing a vibrant and inclusive digital space that benefits individual users and the broader community.

Linking the interpretation of toxicity score data with sentiment classification using VADER and TextBlob provides a comprehensive understanding of digital content dynamics. The toxicity scores highlight the prevalence of harmful language, while sentiment classification reveals the overall emotional tone of the content. For instance, VADER and TextBlob may identify a significant portion of posts as positive. Yet, high toxicity scores indicate that some interactions are negatively impactful despite the generally positive sentiment. This dual analysis underscores the complexity of online behavior, where positive sentiments can coexist with toxic elements. Addressing both toxicity and sentiment is crucial for creating a healthier digital environment. Thus, combining these metrics offers valuable insights into user interactions guiding strategies to enhance community engagement and mitigate negative behaviors.

	# of Posts	Negative Sentiment [-1..-0.05]	Neutral Sentiment (-0.05..0.05)	Positive Sentiment [0.05..1]
VADER (English/EN)	961	315 (32.78%)	165 (17.17%)	481 (50.05%)
TextBlob (English/EN)	961	164 (17.07%)	293 (30.49%)	504 (52.45%)
TextBlob (French/FR)	2	0 (0.00%)	1 (50.00%)	1 (50.00%)
TextBlob (German/DE)	5	0 (0.00%)	5 (100.00%)	0 (0.00%)

Figure 12. Sentiment Classification Result (Based on Vader and TextBlob)

Figure 12 sentiment classification result based on Vader and TextBlob. Based on the analysis of 968 out of 1,021 posts, the sentiment classification using VADER and TextBlob reveals distinct patterns in user sentiment. VADER's analysis of 961 English posts indicates that 315 (32.78%) exhibit negative sentiment, 165 (17.17%) are neutral, and 481 (50.05%) are positive. In comparison, TextBlob classifies 164 (17.07%) posts as unfavorable, 293 (30.49%) as neutral, and 504 (52.45%) as positive. For French and German posts analyzed by TextBlob, all French posts are evenly split between neutral and positive sentiments, while all German posts are categorized as neutral. These results highlight the differences in sentiment detection between the two tools, with VADER identifying a higher proportion of negative sentiment than TextBlob. The variance underscores the importance of using multiple tools for a comprehensive sentiment analysis. Ultimately, these findings provide valuable insights into the emotional landscape of digital content, informing strategies to enhance user engagement and community interaction.

Interpreting the sentiment analysis results and linking them with birder or viewer engagement reveals significant insights into community dynamics. The classification shows a substantial proportion of positive sentiment, particularly with TextBlob indicating 52.45% positive posts, suggesting that the overall atmosphere within the birdwatching community is generally favorable. However, the notable percentage of negative sentiment, especially with VADER identifying 32.78% of negative posts, indicates areas of concern that could affect user engagement and satisfaction. Addressing these negative sentiments is crucial for maintaining a welcoming environment that encourages active participation and retention of users. Consequently, fostering positive interactions and mitigating negativity will enhance the user experience, promoting sustained engagement and growth within the birdwatching community.

The limitations in data processing present several challenges that impact the overall analysis. First, the accuracy of sentiment and toxicity classification is contingent upon the effectiveness of algorithms such as VADER and TextBlob, which may have inherent biases and limitations in handling nuanced language. Additionally, the quality of the initial data is crucial, as noise and irrelevant information can skew the results despite thorough cleaning processes. Another significant limitation is the potential for context loss when isolating text from its broader conversational background, which can lead to misinterpretations. Addressing these limitations requires continuous refinement of data processing techniques and incorporating more sophisticated algorithms to enhance accuracy and reliability. Acknowledging and mitigating these constraints is essential for ensuring robust and meaningful analytical outcomes.

Based on the performance evaluation of the Support Vector Machine (SVM) using SMOTE, the results indicate an accuracy of 76.45% with a standard deviation of ±4.17%. The confusion matrix reveals 356 true negatives, 101 false positives, 126 false negatives, and 381 true positives. The optimistic AUC is 0.821 with a micro average of 0.821, the overall AUC is 0.799, and the pessimistic AUC is 0.776, with similar standard deviations. Precision stands at 75.53% with a micro average of 75.15%, reflecting the model's ability to identify positive instances accurately. The recall is 79.07%, with a micro average of 79.05%, indicating the model's effectiveness in capturing positive cases. The F-measure is 77.04% with a micro average of 77.05%, balancing precision and recall. These metrics demonstrate that the SVM model, enhanced with SMOTE, performs robustly in classifying the data, though there remains room for improvement in handling false positives and negatives.

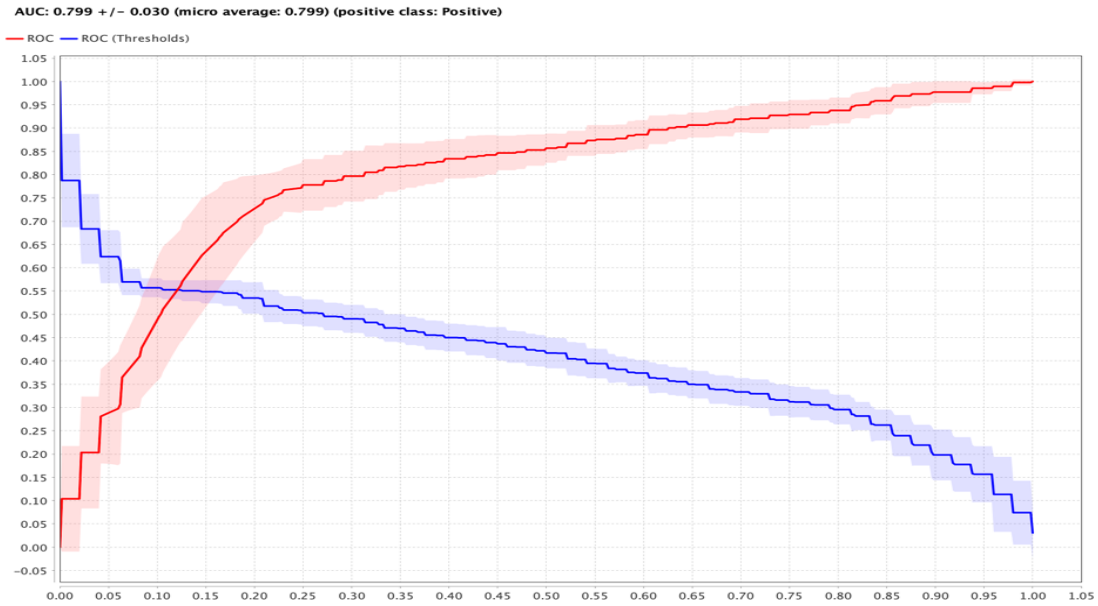

Figure 13. Area Under Curve of SVM with SMOTE in Sentiment Classification

Figure 13 shows the AUC of SVM with SMOTE in Sentiment Classification. Interpreting the AUC of the SVM with SMOTE in sentiment classification reveals the model's performance in distinguishing between positive and negative sentiments. An AUC (Area Under the Curve) of 0.799, with a micro average of 0.799 and a standard deviation of ±0.030, indicates a relatively strong ability of the model to classify sentiment correctly. The ROC (Receiver Operating Characteristic) curve further illustrates the trade-offs between sensitivity and specificity across various threshold settings. The closer the AUC value is to 1, the better the model's performance; thus, an AUC of 0.799 suggests that the SVM with SMOTE performs well, but there is still room for improvement. Enhancing the model to reduce false positives and negatives could increase the AUC, leading to more accurate sentiment classification. Overall, the results demonstrate a robust classification capability, providing a solid foundation for further refinement and application in sentiment analysis tasks.

Analysis based on digital content reviews through toxicity and sentiment classification provides valuable insights into the online community's dynamics and interactions. By evaluating toxicity scores, one can identify the prevalence of harmful language and hostile behavior, directly impacting user experience and engagement. Simultaneously, sentiment classification helps in understanding the overall emotional tone of the content, whether positive, negative, or neutral. This dual approach highlights the relationship between user sentiment and toxic interactions, revealing areas that need attention to foster a healthier and more inclusive environment. Consequently, these analyses inform strategies for moderating content and enhancing user engagement, ultimately contributing to a more positive digital ecosystem.

3.2 Discussion

Digital content and viewer engagement play a crucial role in conveying the content creator's message to the audience. Effective engagement ensures that the viewers consume and interact with the content, enhancing their understanding and retention of the information presented. In the context of birdwatching, well-crafted digital content can educate and inspire viewers about various aspects of the activity, from bird identification to conservation efforts. High levels of viewer engagement can lead to increased interest and participation in birdwatching, fostering a community of enthusiasts who share knowledge and experiences. Thus, strategic content creation and active viewer engagement are essential for maximizing the impact of birdwatching-related digital content.

The connection between tourism and birdwatching is significant, as birdwatching serves as a niche segment within the broader tourism industry. Birdwatching tourism attracts enthusiasts who travel specifically to observe bird species in their natural habitats, contributing to local economies through spending on accommodation, guided tours, and related services [3], [25]. Additionally, birdwatching promotes environmental conservation, raising awareness about preserving natural habitats for avian species. This symbiotic relationship between tourism and birdwatching underscores the potential for sustainable tourism development that benefits both the environment and local communities. Consequently, integrating birdwatching into tourism strategies enhances ecological preservation while driving economic growth.

The analysis of the relationship between digital content and tourism activities, particularly specialized birdwatching tourism, reveals significant insights into how online platforms can enhance niche travel experiences. Digital content, such as informative videos, blogs, and social media posts, is vital in educating and engaging potential birdwatchers, providing them with valuable information about bird species, habitats, and best practices. This content raises awareness and motivates enthusiasts to explore birdwatching destinations, thereby driving tourism in those areas.

Additionally, interactive digital platforms enable birdwatchers to share experiences and connect with like-minded individuals, fostering a sense of community. Thus, effectively leveraging digital content can significantly boost interest and participation in birdwatching tourism, promoting environmental conservation and local economic development.

Digital content related to tourism activities, mainly specialized birdwatching tourism, plays a crucial role in education and participation [26], [26]–[28]. Through various online platforms, content creators provide educational resources that inform viewers about bird species, their habitats, and the ecological significance of birdwatching. This educational aspect enhances enthusiasts' knowledge and fosters a deeper appreciation for environmental conservation. Moreover, interactive digital content encourages active participation by allowing birdwatchers to share their experiences, engage in discussions, and join virtual communities. Consequently, digital content is a powerful tool to promote education and engagement in birdwatching tourism, ultimately contributing to the growth and sustainability of this niche tourism sector.

Connecting digital content in tourism, particularly birdwatching, with toxicity and sentiment analysis provides a comprehensive understanding of online community dynamics. Toxicity analysis identifies negative interactions and harmful language within digital platforms, which can deter user engagement and participation [29], [30]. In contrast, sentiment analysis assesses the emotional tone of the content, revealing how users perceive and interact with birdwatching-related posts. By combining these analyses, stakeholders can identify and address areas of concern, fostering a more positive and inclusive digital environment. This dual approach enhances the effectiveness of digital content as a tool for education and engagement, ensuring that the online birdwatching community remains supportive and welcoming.

In conclusion, integrating digital content into specialized tourism activities, such as birdwatching, significantly enhances educational outreach and participant engagement. Toxicity and sentiment analysis are crucial in understanding and improving the digital community's dynamics, identifying negative interactions, and fostering positive user experiences. These analytical tools provide insights that inform strategies for creating a more supportive and inclusive environment. Ultimately, leveraging digital content effectively promotes environmental awareness and conservation and drives the growth and sustainability of niche tourism sectors. This comprehensive approach ensures that birdwatching enthusiasts can engage in a meaningful and enriching online experience.

4. CONCLUSION

In conclusion, the research demonstrates that integrating digital content into specialized tourism activities, such as birdwatching, significantly enhances educational outreach and participant engagement. The analysis of 1,021 posts reveals an average toxicity score of 0.13839, with VADER identifying 32.78% negative sentiment and TextBlob identifying 17.07% negative sentiment. These findings highlight the importance of addressing negative interactions to maintain a positive online environment. Furthermore, sentiment analysis indicates that over 50% of the posts convey positive sentiment, underscoring the potential of digital content to foster a supportive community. By leveraging these insights, stakeholders can develop targeted strategies to promote environmental awareness and conservation, driving the growth and sustainability of niche tourism sectors. This research underscores the critical role of digital platforms in enhancing user experience and engagement in birdwatching activities.

ACKNOWLEDGMENT

I want to express my sincere gratitude to the Tourism Department, Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, PUSDIPAR, and the LPPM (*Lembaga Penelitian dan Pengabdian kepada Masyarakat*).

REFERENCES

[1] C. Randler and M. Dutour, "Birding specialization and satisfaction in Australian birders–a Big Year is not a big issue," *World Leis. J.*, pp. 1–16, 2024, doi: 10.1080/16078055.2024.2382439.

[2] L. A. Vallejo-Novoa, "Critical mass: the creation of Pajarero/Birder communities in Mexico for citizen science," *Tapuya Lat. Am. Sci. Technol. Soc.*, vol. 6, no. 1, 2023, doi: 10.1080/25729861.2023.2254620.

[3] C. Randler, "Progression through time: Development of birdwatcher careers based on propensity score matching," *Cogent Soc. Sci.*, vol. 9, no. 1, 2023, doi: 10.1080/23311886.2023.2178103.

[4] T. de Oliveira Pinto, "Songbird and birdsong: Listening to the finches in the Harz region, Germany," *Sound Stud.*, vol. 6, no. 2, pp. 215–238, 2020, doi: 10.1080/20551940.2020.1799543.

[5] M. Brooks *et al.*, "The African Bird Atlas Project: a description of the project and BirdMap data-collection protocol," *Ostrich*, vol. 93, no. 4, pp. 223–232, 2022, doi: 10.2989/00306525.2022.2125097.

[6] T. Härtel, J. Vanhöfen, N. Groβmann, and C. Randler, "Unlocking biodiversity awareness: influential factors on bird species knowledge and the links with environmental attitudes and connectedness to nature.," *Int. J. Sci. Educ. Part B Commun. Public Engagem.*, pp. 1–16, 2024, doi: 10.1080/21548455.2024.2381840.

[7] H. D. Oschadleus, "101 Curious Tales of East African Birds: A Brief Introduction to Tropical Ornithology101 Curious Tales of East African Birds: A Brief Introduction to Tropical Ornithology, Author: C Beale, 2023, London: Pelagic Publishing, Paperback GBP 25.00, ebook GBP 20.00, ISBN 9781784272913, 222 pp, https://pelagicpublishing.com/products/101-curious-tales-of-east-

african-birds," *Ostrich*, vol. 94, no. 3, pp. 239–240, 2023, doi: 10.2989/00306525.2023.2274725.
[8] G. Grilli, S. Ferrini, T. Luisetti, and R. Kerry Turner, "The role of choice experiments in natural capital accounting approaches: fast track versus simulated exchange value in the Deben Estuary saltmarshes," *J. Environ. Plan. Manag.*, vol. 65, no. 7, pp. 1281–1300, 2022, doi: 10.1080/09640568.2021.1957794.
[9] D. A. Potvin, B. L. Gilby, M. K. Anderson, and N. I. M. Watson, "Effects of temporal variations in ecotourist noise on an avian community: a case study from a UNESCO world heritage site," *J. Ecotourism*, pp. 1–19, 2023, doi: 10.1080/14724049.2023.2202361.
[10] D. Couceiro, I. R. Hristova, V. Tassone, A. Wals, and C. Gómez, "Exploring environmental stewardship among youth from a high-biodiverse region in Colombia," *J. Environ. Educ.*, vol. 54, no. 5, pp. 306–324, 2023, doi: 10.1080/00958964.2023.2238649.
[11] Å. Davidsson, "Can historic natural disturbances enable conservation opportunities? Evidence from the establishment of national parks in Sweden," *Environ. Hazards*, vol. 23, no. 4, pp. 307–327, 2023, doi: 10.1080/17477891.2023.2282613.
[12] T. C. Young and P. Rautio, "Bewildering the legacy effects of Gail Melson's wild things/animals/children," *Pedagog. Cult. Soc.*, vol. 32, no. 4, pp. 985–1001, 2024, doi: 10.1080/14681366.2024.2355094.
[13] L. J. Forristal, "Students as Co-Creators of Interdisciplinary Tourism Content: A Strategy to Help Prepare Creative, Problem-solving, Research Savvy, and Globally-Competent Hospitality Employees," *J. Hosp. Tour. Educ.*, vol. 35, no. 3, pp. 237–250, 2023, doi: 10.1080/10963758.2022.2034117.
[14] L. J. Backstrom, N. P. Leseberg, C. T. Callaghan, C. Sanderson, R. A. Fuller, and J. E. M. Watson, "Using citizen science to identify Australia's least known birds and inform conservation action," *Emu*, vol. 124, no. 2, pp. 199–205, 2024, doi: 10.1080/01584197.2023.2283443.
[15] V. E. Luzuriaga-Aveiga and D. F. Cisneros-Heredia, "Seasonal turnover of avian community assembly in a highly fragmented Tumbesian dry forest of southwestern Ecuador," *Neotrop. Biodivers.*, vol. 8, no. 1, pp. 229–241, 2022, doi: 10.1080/23766808.2022.2076784.
[16] R. Barcan and J. Johnston, "Birds as buddies: the politics of sentimentality in the Birds in backyards (Australia) Facebook site," *Continuum (N. Y).*, vol. 38, no. 1, pp. 98–110, 2024, doi: 10.1080/10304312.2023.2234106.
[17] S. T. Osinubi, "African–Eurasian birds: our birds, their birds, or OUR birds?," *Ostrich*, vol. 93, no. 3, pp. iii–iv, 2022, doi: 10.2989/00306525.2022.2140518.
[18] I. Literák, V. Reháková, S. Xirouchakis, J. Škrábal, and V. Starenko, "Black kites wintering in Europe: estimated number, subspecies status, and behaviour of a bird wintering on Crete and Turkey," *Eur. Zool. J.*, vol. 89, no. 1, pp. 1271–1284, 2022, doi: 10.1080/24750263.2022.2137253.
[19] P. Pintassilgo, P. Pinto, A. Costa, A. Matias, and M. H. Guimarães, "Environmental attitudes and behaviour of birdwatchers: a missing link," *Tour. Recreat. Res.*, vol. 48, no. 3, pp. 399–418, 2023, doi: 10.1080/02508281.2021.1920755.
[20] A. Giampiccoli, D. O. Mtapuria, and S. Jugmohan, "Community-based tourism and animals: Theorising the relationship," *Cogent Soc. Sci.*, vol. 6, no. 1, 2020, doi: 10.1080/23311886.2020.1778965.
[21] A. Chakraborty and T. Takenaka, "Tourism and sustainability at a crossroads in Shiretoko UNESCO World Natural Heritage Site in Japan," *J. Herit. Tour.*, vol. 18, no. 2, pp. 202–223, 2023, doi: 10.1080/1743873X.2023.2169614.
[22] C. Hehir, C. Scarles, K. J. Wyles, and J. Kantenbacher, "Last chance for wildlife: making tourism count for conservation," *J. Sustain. Tour.*, vol. 31, no. 5, pp. 1271–1291, 2023, doi: 10.1080/09669582.2022.2049804.
[23] I. Mertena and M. Kaaristo, "Understanding train tourism mobilities : a practice theories perspective perspective," *Mobilities*, vol. 0, no. 0, pp. 1–17, 2024, doi: 10.1080/17450101.2024.2316827.
[24] Y. A. Singgalen, "Analysis and design of Hutanku application using Rapid application development," *Mantik J.*, vol. 7, no. 4, pp. 3461–3470, 2024.
[25] D. K. M., N. S. Govindarajo, and M. H. S. Khen, "Effect of service quality on visitor satisfaction, destination image and destination loyalty – practical, theoretical and policy implications to avitourism," *Int. J. Cult. Tour. Hosp. Res.*, vol. 14, no. 1, pp. 83–101, Jan. 2020, doi: 10.1108/IJCTHR-04-2019-0066.
[26] E. Sánchez-Amboage, M. Enrique Membiela-Pollán, V. A. Martínez-Fernández, and S. Molinillo, "Tourism marketing in a metaverse context: the new reality of European museums on meta," *Museum Manag. Curatorsh.*, vol. 38, no. 4, pp. 468–489, 2023, doi: 10.1080/09647775.2023.2209841.
[27] I. C. Dewi, M. Soediro, and M. R. Sondak, "Peluang Digital Marketing Dalam Kebangkitan Pariwisata Di Indonesia," *JMBI UNSRAT (Jurnal Ilm. Manaj. Bisnis dan Inov. Univ. Sam Ratulangi)*, vol. 9, no. 2, pp. 969–974, 2022, doi: 10.35794/jmbi.v9i2.43938.
[28] K. H. Mkwizu, "Digital marketing and tourism: opportunities for Africa," *Int. Hosp. Rev.*, vol. 34, no. 1, pp. 5–12, Jan. 2020, doi: 10.1108/IHR-09-2019-0015.
[29] Y. A. Singgalen, "Toxicity Analysis and Sentiment Classification of Wonderland Indonesia by Alffy Rev using Support Vector Machine," *J. Sist. Komput. dan Inform.*, vol. 5, no. 3, pp. 538–548, 2024, doi: 10.30865/json.v5i3.7563.
[30] Y. A. Singgalen, "Toxicity , topic , and sentiment analysis on the operation of coal-fired power plants content reviews," *J. Tek. Inform. C.I.T Medicom*, vol. 16, no. 1, pp. 45–57, 2024.