Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

Analyzing Social Networks and Topic Clustering in Backpacker Tourism Content Reviews using K-means, Fast HDBScan, and Gaussian Mixture with Communalytic

Yerik Afrianto Singgalen

Faculty of Business Administration and Communication, Tourism Study Program, Atma Jaya Catholic University of Indonesia, Jakarta
Jl. Jend. Sudirman No.51 5, RT.004/RW.4, Karet Semanggi, Setiabudi District, South Jakarta City, Special Capital Region of Jakarta, Indonesia
Email: yerik.afrianto@atmajaya.ac.id
Correspondence Author Email: yerik.afrianto@atmajaya.ac.id
Submitted: **25/09/2024;** Accepted: **09/10/2024;** Published: **13/10/2024**

Abstract−This research explores the integration of social network analysis and topic clustering techniques to provide novel insights into digital interactions and thematic trends within the context of backpacker tourism. Utilizing a structured framework, 3,575 records across three content IDs (c2ZMFDS_3rU, Sv_yxz7T8rU, and i9t9pbdo-bk) were processed and classified into 10 clusters using k-Means, Fast HDBScan, and Gaussian Mixture algorithms. Social network analysis was performed on 4,224 actor nodes and 395 edges, highlighting the role of key influencers in driving conversations while revealing the participation patterns of a larger, less engaged audience. The topic clustering revealed distinct themes, including budget travel, off-the-beaten-path destinations, and sustainable tourism, with each algorithm offering unique insights into the structure of the data. The novelty of this research lies in applying these computational methods to backpacker tourism, traditionally analyzed through qualitative approaches, to uncover how thematic discussions propagate within digital communities. By integrating these techniques, the study provides a deeper understanding of how key topics resonate with backpackers and how social interactions influence the spread of ideas. The findings offer valuable implications for content creators and tourism marketers seeking to engage this niche travel demographic more effectively. This work contributes a scalable, data-driven methodology for analyzing traveler behavior and preferences in virtual environments, enhancing the field of backpacker tourism research.

Keywords: Social Network; Topic Clustering; K-Means; Fast HDBScan; Gaussian Mixture

1. INTRODUCTION

Mapping and analyzing social network viewers' responses to backpacker tourism content holds significant value in identifying the forms and patterns of social networks in virtual spaces. The dynamic nature of these digital interactions reflects diverse audience engagements, where individuals participate actively in content consumption, sharing, and feedback [1]. Each response, whether expressed through likes, comments, or shares, represents a critical node within the broader digital network, offering insights into user behavior and influence within specific communities [2], [3]. By systematically examining these patterns, a clearer understanding emerges of how virtual communities form, evolve, and interact with content, especially in the context of niche interests such as backpacker tourism [4]–[6]. This process of analysis not only reveals the immediate reach and impact of the content but also the underlying structures that govern social connections and influence [7]–[9]. Identifying key influencers and understanding the flow of information within these networks becomes essential for comprehending the social architecture of virtual spaces. Ultimately, this investigation contributes to a more refined grasp of how digital platforms facilitate interconnectedness and community dynamics around shared experiences in the travel domain.

Topic clustering plays a crucial role in categorizing viewer comments related to backpacker tourism content, facilitating the identification of traveler preferences regarding suitable backpacking activities. Through this method, similar comments are grouped based on shared themes or ideas, allowing for a more structured and coherent understanding of audience feedback [10], [11]. The process enables a deeper analysis of prevalent opinions, uncovering patterns that reflect specific interests or concerns within the backpacker community [12]–[15]. This clustering method offers the advantage of organizing vast amounts of unstructured data into meaningful segments, which helps highlight the dominant trends in tourism behavior [16]–[19]. By examining these clusters, a clearer picture of the most appealing destinations, activities, or experiences in the realm of backpacking emerges, offering valuable insights for tourism providers aiming to align services with traveler expectations. Thus, this analytical approach contributes to more targeted strategies in content creation and service offerings, enhancing the overall tourism experience.

The objective of research on social network and topic clustering of backpacker tourism content reviews is to uncover patterns of interaction and thematic preferences within digital communities. This investigation seeks to map the social structures that emerge from user engagement with backpacker tourism content, while simultaneously categorizing and analyzing the topics that dominate these discussions [20]. The application of the K-means, Fast HDBScan, and Gaussian Mixture algorithm serves to efficiently organize the vast and diverse data generated by user reviews, enabling a systematic identification of clusters that reflect common interests and behaviors [21]. By isolating these clusters, the study offers a nuanced understanding of how information flows within virtual social networks and what themes resonate most with backpackers. The integration of social network

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

analysis with topic clustering enhances the accuracy of insights into audience behavior, thus contributing to more effective content strategies and tourism services that are better aligned with traveler needs and expectations.

The urgency of this research lies in its capacity to address the growing complexities of user-generated content and the evolving dynamics of virtual communities, particularly within the context of backpacker tourism. As digital platforms increasingly shape travelers' decisions and interactions, understanding the underlying social and thematic patterns becomes essential [22]. The rapid proliferation of online reviews and discussions necessitates advanced methods of data analysis, such as clustering algorithms, to systematically process and interpret large-scale information. Failing to examine these trends risks overlooking critical insights into traveler preferences and behaviors, which are pivotal for shaping effective tourism strategies [23]. By identifying key influencers, recurring themes, and patterns of engagement, this research serves as a vital tool for stakeholders aiming to enhance their alignment with contemporary traveler expectations. Consequently, the study not only contributes to academic discourse but also holds practical relevance for industries navigating the digital transformation of tourism.

The theoretical and practical contributions of this research are integral to advancing both academic understanding and real-world applications within the field of backpacker tourism. Theoretically, the study enriches the literature on social network analysis and topic clustering by offering a nuanced examination of how virtual interactions and thematic preferences shape user-generated content [24]. This approach deepens insights into the structure and behavior of digital communities, contributing to a broader comprehension of online engagement patterns [25]. Practically, the research provides valuable tools for tourism professionals seeking to refine their strategies based on the preferences of backpackers. The implementation of clustering algorithms enables more precise targeting of services and content, ensuring a better alignment with the interests and expectations of this particular travel demographic [26], [27]. These findings hold significant relevance for both digital marketing and tourism management, as they support more informed decision-making processes. Overall, the research bridges the gap between theoretical exploration and actionable insights, offering benefits for both academic inquiry and industry practice.

Similar research in the field of social network and topic analysis, particularly in the context of tourism, has focused on exploring the ways in which user-generated content influences travel behavior and decision-making processes. Studies have employed various clustering algorithms and network analysis techniques to examine patterns of engagement and thematic trends across different types of tourism content, including adventure, cultural, and eco-tourism [28]–[30]. These works highlight the growing importance of understanding digital interactions, as they reveal how virtual communities shape perceptions of destinations and travel experiences. By analyzing user reviews, comments, and social media posts, prior research has contributed to identifying key influencers, preferences, and behaviors within specific traveler groups [31]–[33]. Despite the progress made, there remains a need for further investigation into the niche domain of backpacker tourism, where social dynamics and preferences may differ significantly from other types of travelers. This gap in the literature underscores the relevance of applying advanced methods, such as the K-means clustering algorithm, to gain deeper insights into the unique characteristics of backpacker networks and content preferences, thus advancing both theoretical understanding and practical applications in this area.

The limitations of this research are primarily linked to the inherent challenges of analyzing large-scale, user-generated content within the digital landscape of backpacker tourism. One significant constraint arises from the variability and subjectivity of online reviews and social media interactions, which can lead to inconsistencies in data quality and relevance. Furthermore, the application of the K-means clustering algorithm, while effective in categorizing topics, may oversimplify complex social interactions and fail to account for subtler nuances within user behavior. The algorithm's reliance on predefined clusters could also result in the exclusion of emerging or less frequent topics that might still hold valuable insights. Additionally, the dynamic and rapidly changing nature of online platforms introduces a temporal limitation, as the data analyzed may quickly become outdated due to shifts in user preferences or platform algorithms. These limitations suggest that while the research offers meaningful contributions, its findings must be interpreted within the context of these constraints, acknowledging that further refinement and complementary approaches may be needed to fully capture the complexities of digital backpacker tourism networks.

2. RESEARCH METHODOLOGY

2.1 Social Network and Topic Clustering in Tourism

Social networks in tourism play a pivotal role in shaping travel behavior, facilitating the exchange of information, and fostering community engagement among travelers. These digital platforms allow individuals to share experiences, reviews, and recommendations, thus influencing others' perceptions of destinations and tourism services [34]. The interconnectedness fostered by social networks creates a dynamic space where personal narratives and peer reviews often carry more weight than traditional advertising, significantly impacting travel decisions [35]. This phenomenon highlights the growing reliance on user-generated content as a primary source of information for tourists seeking authentic experiences. Analyzing these networks provides valuable insights into how travelers form opinions, interact with destinations, and influence one another within virtual communities.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

Consequently, understanding social network dynamics is crucial for tourism providers aiming to tailor their services to meet the evolving needs and preferences of modern travelers, underscoring the transformative impact of digital interactions on the tourism industry.

Topic clustering in tourism serves as an essential method for organizing and analyzing vast amounts of unstructured data generated by online platforms, particularly in the form of user reviews, comments, and social media interactions. By grouping similar topics together, this approach allows for a clearer understanding of common themes and preferences among travelers, shedding light on what drives their decisions and how they engage with tourism-related content [36]. The ability to identify clusters of interest not only reveals insights into destination popularity or preferred activities but also highlights emerging trends and niche markets that might otherwise go unnoticed [37]. This structured analysis is particularly valuable in enhancing the accuracy of marketing strategies and improving service offerings, as it enables tourism providers to better align their offerings with the specific needs and desires of their target audiences. Ultimately, topic clustering contributes to a more efficient and targeted approach in the tourism sector, facilitating a deeper connection between travelers and the experiences they seek.

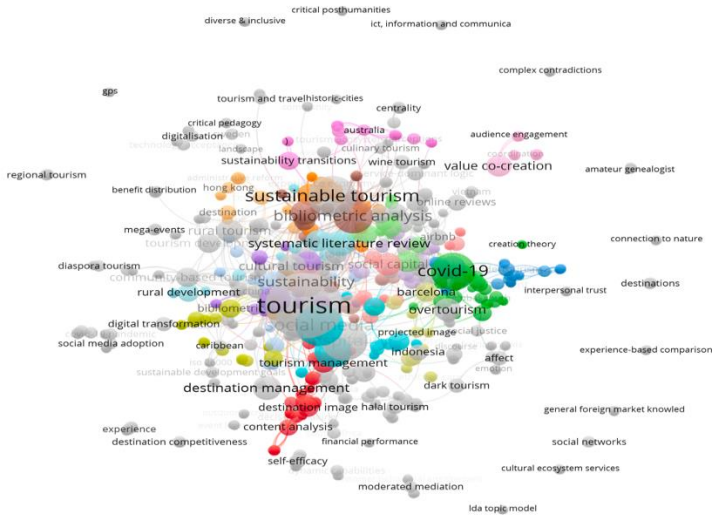

Figure 1. Social Network and Topic Clustering in Vrious Tourism

Figure 1 presents a visualization of social network and topic clustering in various aspects of tourism, effectively mapping the connections between keywords and thematic areas within the domain. At the core of the diagram, larger nodes such as “tourism” and “sustainable tourism” indicate central topics that dominate the discourse, with closely related terms forming clusters around them. These clusters represent interconnected topics that share common ground in the broader conversation, providing insight into how certain themes like “destination management” or “COVID-19” influence discussions within the tourism sector. The varying sizes and colors of the nodes suggest different levels of engagement and thematic relevance, where larger and more vibrant clusters highlight areas of intense focus. This visual representation not only helps in understanding the complexity of interactions within the tourism network but also identifies emerging trends and key topics that are shaping current discussions. Such clustering aids in recognizing the diversity of themes and the potential interrelations between various facets of tourism, offering valuable insights for both academic research and practical applications in industry planning.

The novelty of this research lies in its integration of social network analysis with advanced topic clustering techniques to explore the digital landscape of tourism discourse, particularly within the niche of backpacker tourism. This approach distinguishes itself by not only examining the content generated by users but also by mapping the relational structures that underpin these discussions in virtual spaces. The dual focus on both thematic clustering and network dynamics offers a more comprehensive understanding of how ideas and preferences propagate through online communities. This method presents a unique contribution to tourism studies by providing actionable insights into traveler behavior and preferences, using a data-driven approach that uncovers patterns that may not be immediately apparent through traditional qualitative analyses. Consequently, this research introduces a new dimension to the exploration of digital interactions in tourism, paving the way for future studies to build upon its findings and methodologies, while also offering practical implications for industry stakeholders seeking to harness the power of social networks in shaping tourism experiences.

2.2 Digital Content Reviews and Analysis Framewok

The framework employed in this research, known as the Digital Content Reviews and Analysis Framework, provides a structured approach to processing and analyzing large volumes of user-generated data within the tourism

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34–48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

sector. This framework systematically organizes and filters diverse forms of digital content, such as reviews, comments, and social media interactions, to extract relevant insights regarding traveler preferences and behaviors [38]–[40]. Its robust design allows for the efficient categorization of unstructured data into meaningful clusters, enabling a deeper understanding of the themes and patterns that shape discussions in virtual communities [41], [42]. By incorporating both qualitative and quantitative elements, this framework offers a comprehensive tool for examining the intricate relationships between content and user engagement. The use of this framework not only enhances the precision of the data analysis process but also ensures that findings are relevant and applicable to practical contexts, providing valuable insights for both academic research and tourism industry stakeholders. This approach contributes to a more informed decision-making process, aligning tourism services with the evolving demands of modern travelers.

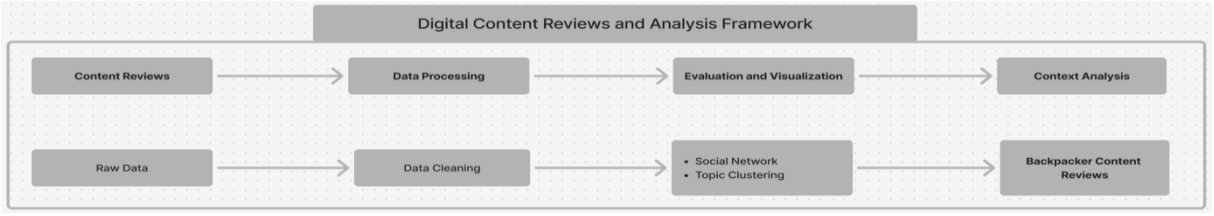

Figure 2. Digital Content Reviews and Analysis Framework

Figure 2 illustrates the Digital Content Reviews and Analysis Framework, a structured approach designed to process and analyze user-generated reviews in the context of backpacker tourism. The framework begins with the collection of raw data, followed by the essential steps of data cleaning, selection, and extraction to ensure the relevance and quality of the information. Once processed, the data undergoes evaluation and visualization, which helps in assessing model performance and understanding the visualized insights. This phase is critical for identifying patterns and trends within the data, supporting a more refined interpretation of social network dynamics and topic clustering related to backpacker content. The final step in the framework is the content analysis, where the processed data is analyzed in depth to uncover key themes, behaviors, and influences within the virtual tourism community. By systematically organizing the flow from raw data to actionable insights, this framework enables a comprehensive analysis of digital reviews, ultimately contributing to a clearer understanding of user preferences and social interactions in tourism.

The Digital Content Reviews and Analysis Framework is highly relevant to this research as it provides a systematic method for processing and analyzing vast amounts of user-generated content, particularly in the context of backpacker tourism. This framework offers a structured pathway from raw data collection to insightful analysis, ensuring that the data is not only cleansed and refined but also organized in a way that highlights significant patterns and behaviors within the virtual tourism community. By utilizing such a framework, this research gains the ability to more accurately identify key trends, preferences, and social interactions that shape backpacker tourism, offering deeper insights into the behaviors of travelers. Additionally, the integration of social network analysis and topic clustering into this framework enhances its analytical precision, allowing for a more granular understanding of how information is shared and engaged with online. As a result, this framework not only supports the methodological rigor of the study but also ensures the findings are actionable and relevant for both academic inquiry and practical application in the tourism industry.

2.2.1 Content Reviews

In the content review phase, identifying frequently used words from comments based on the videos with IDs: c2ZMFDS_3rU, Sv_yxz7T8rU, and i9t9pbdo-bk is crucial for understanding viewer engagement and thematic relevance. This step allows for a systematic extraction of key terms that reflect the audience's primary interests, concerns, and emotional responses to the content. By analyzing the most common words, patterns of discussion can be detected, providing valuable insights into the preferences and reactions of the viewers. Such analysis also aids in uncovering the dominant themes or recurring topics, which can help content creators align future productions with audience expectations. The frequency of specific terms serves as an indicator of what resonates most within the community, and this lexical analysis contributes to more targeted strategies in both content development and audience engagement, enhancing the overall impact of the digital content in the tourism sector.

Video 1

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

Video 2

Video 3

Figure 3. Frequently used words (Communalytic)

Figure 3 presents the frequently used words from the comments of three different videos, analyzed using Communalytic, which reveals patterns of audience interaction and engagement. In Video 1, terms such as "keren," "bangnya," and "video" appear prominently, indicating a strong focus on the admiration for the video content and the creator, likely reflecting positive viewer feedback. Video 2 similarly highlights words like "keren," "bang," and "subscribe," suggesting a high level of viewer enthusiasm and encouragement for content subscription and continued engagement with the creator. In Video 3, the word cloud continues to emphasize terms like "keren," "video," and "bang," reinforcing the trend of audience appreciation and positive sentiment towards the content. The consistent appearance of these terms across all videos points to a recurring theme of praise, interactivity, and a sense of community within the viewer base. This lexical analysis helps in understanding the emotional tone and behavioral patterns of the audience, providing insight into how viewers perceive and engage with backpacker tourism content. The results indicate strong positive feedback, which can be leveraged for enhancing future content strategies.

The next step involves identifying the emoji cloud from each video to gain deeper insights into the emotional tone and engagement level expressed by the viewers. Emojis serve as non-verbal cues that complement textual comments, providing a visual representation of audience reactions, which can range from positive emotions such as excitement and admiration to more neutral or critical sentiments. By analyzing the frequency and variety of emojis used in the comments, a clearer understanding of the emotional landscape surrounding each video can be obtained. This analysis will help to discern whether the content consistently elicits certain types of emotional responses or whether different videos provoke varying reactions from the audience. Moreover, the emoji cloud offers a more accessible and intuitive way of capturing the general mood and enthusiasm of viewers, which may not always be fully conveyed through words alone. Ultimately, this step enhances the overall content analysis by adding a valuable layer of emotional interpretation, enriching the understanding of viewer engagement and sentiment.

Video 1

Video 2

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

Video 3

Figure 4. Emoji Cloud (Communalytic)

Figure 4 illustrates the emoji cloud generated from viewer comments on three different videos, revealing the emotional responses and engagement of the audience. In Video 1, emojis such as the "thumbs up," "heart eyes," and "fire" dominate, indicating strong positive reactions, admiration, and excitement about the content. Video 2 displays a similar pattern, with "laughing face," "heart eyes," and "thumbs up" appearing frequently, which suggests that viewers found the content both enjoyable and visually appealing. Video 3 continues this trend, with prominent use of emojis like "heart eyes," "thumbs up," and "crying face," signifying a mix of enthusiasm, approval, and perhaps emotional connection or humor. The consistency of these emojis across the three videos points to a general sentiment of positivity and high viewer engagement, reflecting the audience's strong emotional investment in the content. This emoji cloud analysis serves as an effective tool for gauging emotional tone and engagement levels, offering insights that complement the textual feedback, thus enhancing the understanding of audience response to digital content.

Based on the identification of the word and emoji cloud, the text data can be effectively cleaned to enhance the precision and clarity of the analysis. The process of cleaning involves removing redundant words, irrelevant symbols, and repetitive emojis that do not contribute meaningful insights to the overall sentiment or thematic analysis. By filtering out such noise, the dataset becomes more refined, allowing for a clearer focus on the core topics and emotional expressions that are most representative of viewer engagement. This step is essential in ensuring that the analysis remains relevant and accurate, as it eliminates any distortions that could skew the results. Through this methodical data cleaning process, the subsequent analysis will yield more reliable and actionable insights into user behavior and preferences, providing a solid foundation for deeper content analysis and the development of more targeted digital strategies in tourism or other sectors.

2.2.2 Data Processing

The data cleaning process is crucial for optimizing the subsequent steps in clustering or topic analysis, as it ensures that the dataset is free from noise and irrelevant information. By removing inconsistencies, duplicates, and unnecessary tokens, the dataset becomes more structured and focused, allowing for a more accurate and efficient clustering process. Clean data enhances the precision of topic clustering algorithms, as it eliminates the distortion caused by redundant or extraneous elements that might otherwise skew the results. Furthermore, the refined dataset allows for a clearer identification of patterns and relationships between topics, leading to more meaningful insights. This streamlined data provides a solid foundation for the topic clustering process, ensuring that the analysis reflects the true thematic structure of the content. Ultimately, proper data cleaning is an essential step that significantly impacts the reliability and relevance of the final clustering outcomes.

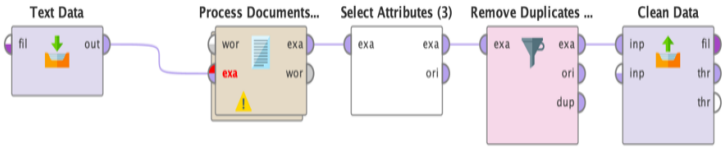

Figure 5. Cleaning Process (Rapidminer)

Figure 5 illustrates the data cleaning process using RapidMiner, which methodically transforms raw text data into a refined dataset ready for further analysis. The workflow begins with the input of text data, which is then processed through various steps, including tokenization and attribute selection. During the "Process Documents" stage, the text is broken down into individual components, allowing for the extraction of relevant features. Next, attributes are selected to ensure that only the most pertinent data is retained, effectively streamlining the dataset. The removal of duplicates follows, eliminating any redundant entries that could distort the analysis. This process not only refines the data but also enhances its integrity, making it more suitable for subsequent tasks such as clustering or sentiment analysis. The final "Clean Data" step outputs a dataset that is both structured and devoid of irrelevant elements, significantly improving the accuracy and reliability of the research findings. This systematic approach ensures that the resulting dataset is optimized for insightful and precise analytical outcomes.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

Figure 6. Proses Documents form Data (Rapidminer)

Figure 6 illustrates the process of document handling in RapidMiner, showcasing the sequential steps involved in transforming raw text data into a structured format suitable for further analysis. The workflow begins with tokenization, where the text is broken down into individual units or tokens, facilitating easier processing and analysis. Following this, the "Transform Case" step ensures consistency by converting all text to a uniform format, such as lowercasing, to avoid duplication of identical terms presented in different cases. Subsequently, irrelevant tokens are filtered out, allowing the retention of meaningful data while discarding unnecessary elements. The "Filter Stopwords" stage further refines the data by removing common but non-informative words like conjunctions and articles that do not contribute significant insights. Finally, the cleaned and organized data is saved, ensuring it is ready for the next phase of analysis, such as clustering or sentiment evaluation. This structured, methodical approach enhances the efficiency of the analytical process and ensures the reliability of the results by focusing only on the most relevant textual components.

Cleaned data is significantly easier to group using the K-means algorithm, as the process ensures that the dataset is free from noise and irrelevant elements, allowing for more accurate clustering. By eliminating redundant information, such as stopwords and inconsistent formatting, the K-means algorithm can focus on the most relevant features of the data, which are essential for identifying distinct clusters. The precision of the algorithm improves because it can now operate on a dataset that clearly represents the underlying patterns and relationships within the content. This streamlined data facilitates the algorithm's ability to assign data points to the nearest centroid, leading to well-defined clusters that reflect meaningful groupings. Consequently, the accuracy of the clustering results is enhanced, making it easier to extract actionable insights from the data, whether for market segmentation, sentiment analysis, or topic modeling. This preparation ultimately maximizes the utility of the K-means algorithm in revealing key patterns within the dataset.

2.2.3 Evaluation and Visualization

In the evaluation phase of topic clustering, a comparative analysis is conducted between the results of data post grouping using K-means, Fast HDBScan, and Gaussian Mixture algorithms. Each algorithm approaches clustering with distinct methodologies, offering varied insights into the structure of the data. K-means, with its centroid-based approach, efficiently creates well-separated clusters, but it may struggle with clusters of irregular shapes. Fast HDBScan, which is density-based, excels in identifying clusters of varying densities and shapes, making it particularly useful for datasets with noise or anomalies. Meanwhile, the Gaussian Mixture model, which assumes that data points are generated from a mixture of several Gaussian distributions, provides a probabilistic clustering method that can capture more complex cluster shapes and overlapping distributions. By comparing the outputs of these algorithms, it becomes possible to assess which method produces the most coherent and meaningful clusters based on the nature of the dataset, ultimately guiding the choice of the most appropriate technique for extracting valuable insights from the data.

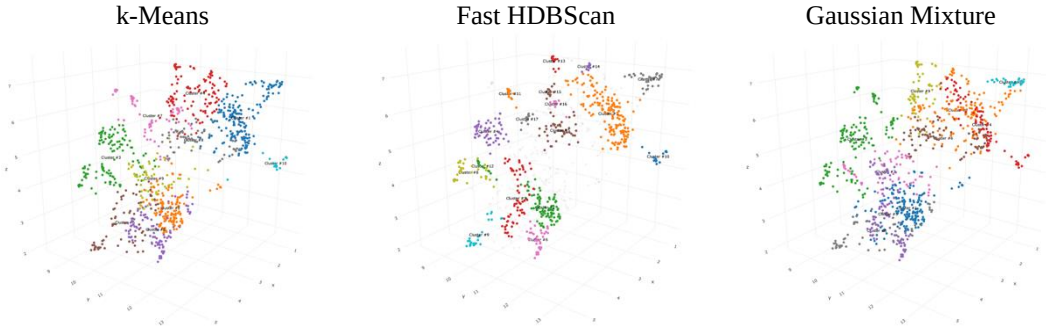

Figure 7. Comparative of K-means, Fast HDBScan, and Gaussian Mixture in Topic Clustering

Figure 7 illustrates the clustering results from three different algorithms K-means, Fast HDBScan, and Gaussian Mixture applied in the modeling process. K-means, displayed in the first visualization, uses a centroid-based approach to form clusters, where data points are assigned to the nearest centroid. This method is efficient but may struggle with clusters of non-spherical shapes. In the second visualization, Fast HDBScan, a density-based clustering method, is shown to excel in identifying clusters with varying densities, making it more robust in handling noise and outliers within the dataset. The third visualization, representing the Gaussian Mixture model, assumes that data points are generated from a mixture of several Gaussian distributions, allowing for the formation of more flexible and overlapping clusters. The Gaussian Mixture method offers a probabilistic approach, giving each data point a probability of belonging to a cluster rather than a strict assignment. By comparing these models,

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

it becomes evident that each algorithm provides unique insights, with K-means offering simplicity and speed, Fast HDBScan handling noise more effectively, and Gaussian Mixture capturing more complex data distributions, guiding the selection of the most suitable algorithm for a given dataset.

In addition to topic clustering, the visualization of social networks is analyzed based on threaded discussions, providing deeper insights into the structure and dynamics of user interactions. Threaded discussions capture the flow of conversations, revealing the relationships between individual comments and responses, which allows for the identification of influential nodes or users within the network. This method highlights how information and opinions propagate within the community, offering a nuanced understanding of how certain topics gain prominence or traction. Analyzing these social networks in conjunction with topic clustering not only uncovers key themes but also elucidates the social architecture behind the discussions, showing which users drive engagement and how sub-communities form around specific topics. Such visualizations enable a more comprehensive understanding of both the content and the social dynamics, allowing for more targeted strategies in managing online discussions or enhancing user engagement in virtual communities.

2.2.4 Context Analysis

At the context analysis stage, post data will be examined within the framework of backpacker tourism, making social network and topic clustering essential for identifying popular topics among viewers. This analysis allows for the extraction of key themes and patterns related to backpacking experiences, as reflected in user interactions and discussions. By combining topic clustering with social network analysis, it becomes possible to understand not only the most discussed topics but also how these themes spread and gain influence within the online community. This dual approach provides a deeper understanding of the digital discourse surrounding backpacker tourism, highlighting the content that resonates most with the audience. Furthermore, this method helps identify influential users and sub-communities, offering a clearer picture of how preferences and trends emerge within these virtual spaces. Ultimately, the integration of social network and topic clustering enriches the context analysis by offering a comprehensive view of both the content and the social dynamics at play.

Understanding the context of backpacker tourism allows post analysis to be aligned with relevant tourism concepts that are specific to backpacking activities. This alignment enables a more focused interpretation of user-generated content, as it takes into account the unique preferences, behaviors, and motivations of backpackers. By integrating these concepts into the analysis, such as budget-conscious travel, adventurous experiences, and cultural immersion, the data becomes more meaningful and directly applicable to the study of backpacker tourism. This approach not only enhances the accuracy of topic identification but also enriches the insights drawn from social network interactions, as it contextualizes the discussions within the framework of relevant tourism theories. As a result, the analysis provides a clearer understanding of how backpacking trends and experiences are discussed, allowing for a more targeted exploration of the factors that influence backpacker decision-making and engagement in online communities.

3. RESULT AND DISCUSSION

The discussion in this research focuses on the exploration of topic clustering and the social network of backpacker tourism content, aiming to uncover the thematic patterns and social dynamics prevalent within this niche. Topic clustering allows for the identification of key themes that emerge in backpacker-related discussions, providing a detailed map of popular subjects such as budget travel, adventure activities, and cultural exchanges. The social network analysis further complements this by mapping the relationships and interactions between content creators, revealing influential nodes and the structure of communication within the backpacker community. This combined approach not only highlights the most prominent topics but also exposes how information flows and clusters within this network. The interaction between content clusters and social connections underscores the role of influential figures in shaping discourse and spreading travel trends. Therefore, this analysis is instrumental in understanding both the thematic landscape of backpacker tourism and the social architecture that supports it, offering insights into the flow of information and influence across the network.

3.1 Topic Clustering and Social Network of Backpacker Tourism Content (c2ZMFDS_3rU)

In the video content with ID c2ZMFDS_3rU, 962 records were classified into 10 clusters using both the k-Means and Gaussian Mixture algorithms, while Fast HDBScan generated 10 clusters with an epsilon value of 0.1. The use of k-Means and Gaussian Mixture provided clear, well-defined clusters based on centroid and probabilistic modeling, respectively, which are effective for organizing structured data. In contrast, Fast HDBScan's density-based approach allowed for the identification of clusters that vary in shape and density, making it particularly useful for managing noise within the dataset. By setting the epsilon parameter to 0.1, Fast HDBScan effectively separated the data into distinct groups, highlighting its robustness in handling less structured data. This comparative clustering process demonstrates the flexibility of each algorithm in capturing the underlying patterns within the content, offering multiple perspectives on how user interactions and comments can be grouped for

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

deeper analysis of engagement trends. Each method provides valuable insights, reinforcing the importance of algorithm selection based on dataset characteristics and analytical objectives.

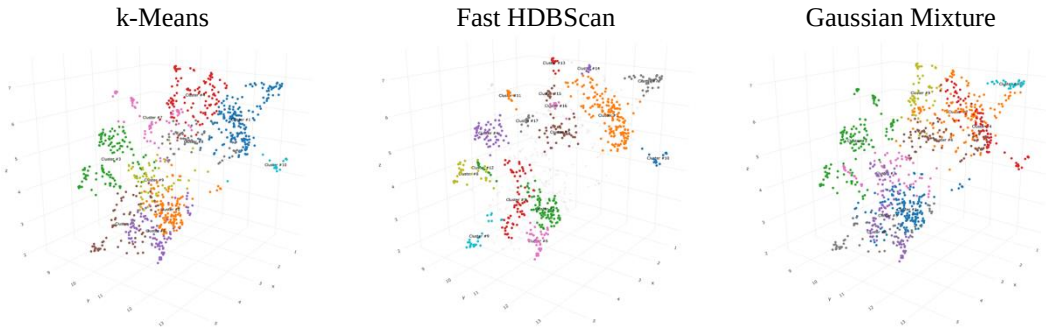

Figure 8. Topic Clustering using k-Means, Fast HDBScan, and Gaussian Mixture (Communalytic)

Figure 8 illustrates the results of topic clustering using three different algorithms: k-Means, Fast HDBScan, and Gaussian Mixture, as visualized through Communalytic. The k-Means clustering, depicted in the first visualization, organizes the data into distinct, non-overlapping groups based on the proximity of data points to centroids, providing clear and well-separated clusters. Fast HDBScan, seen in the second visualization, takes a density-based approach, excelling in identifying clusters of varying densities while handling noise and outliers more effectively than k-Means, which makes it suitable for datasets with complex structures. The Gaussian Mixture model, presented in the third visualization, employs a probabilistic approach, allowing for overlapping clusters and offering a more flexible interpretation of the data's underlying structure. Each method brings unique strengths, with k-Means providing simplicity and speed, Fast HDBScan handling data irregularities, and the Gaussian Mixture capturing nuanced relationships in overlapping clusters. This comparison highlights the importance of selecting the appropriate algorithm based on the characteristics of the dataset and the desired outcomes of the analysis.

In addition, the network is composed of 836 actor nodes connected by 143 edges, reflecting the structure and interaction patterns within the digital community. The actor nodes represent individual participants engaging with the content, while the edges symbolize the relationships formed through their interactions, such as comments, replies, or shared content. This network configuration reveals the degree of connectivity and influence among users, with some nodes acting as central hubs that facilitate communication and others representing more peripheral participants. The relatively low number of edges compared to nodes suggests a dispersed network where interactions are limited to certain key actors rather than widespread engagement across all users. This analysis provides insight into the dynamics of the social network, illustrating how information or influence is disseminated within the group and identifying potential influencers or sub-communities within the broader network. Such findings are essential for understanding how digital content fosters engagement and community-building within virtual spaces.

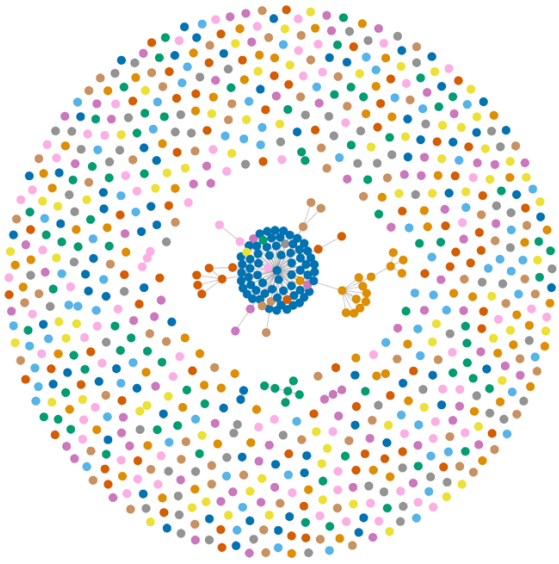

Figure 9. Threaded Discussion Network based on Content id c2ZMFDS_3rU (Communalytic)

Figure 9 illustrates the threaded discussion network based on content ID c2ZMFDS_3rU, visualized through Communalytic. This network shows the interactions between participants, with nodes representing

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

individual actors and edges signifying the connections formed through direct replies and threaded discussions. The dense central cluster indicates a core group of highly engaged users who actively participate in the conversation, while the more dispersed outer nodes reflect participants with fewer interactions or less frequent involvement. The visual structure highlights the hierarchical nature of the discussion, where the central nodes likely serve as key influencers, driving the dialogue and shaping the flow of information. Meanwhile, the peripheral nodes suggest a broader audience that interacts less but remains connected to the core discussion. This network visualization is essential for understanding the dynamics of community engagement, identifying influential users, and analyzing how discussions propagate through different levels of participation within the digital space.

The results of the social network analysis and topic clustering for content ID c2ZMFDS_3rU provide a comprehensive understanding of both user interactions and thematic discussions within the digital community. The social network analysis reveals a dense core of active participants, surrounded by a more dispersed group of users with lower engagement levels, indicating that a small number of key actors drive much of the conversation. These influential users play a central role in shaping the discussion and facilitating the flow of information. The topic clustering, performed using algorithms like k-Means, Fast HDBScan, and Gaussian Mixture, further identifies distinct themes that resonate with the audience. Each clustering method highlights different aspects of the data, with k-Means emphasizing clear, non-overlapping groups, while Fast HDBScan uncovers more complex, density-based structures, and Gaussian Mixture captures overlapping discussions. Together, these analyses offer valuable insights into both the structure of the conversation and the topics that generate the most engagement, underscoring the interconnected nature of content themes and user dynamics in shaping the overall digital discourse.

3.2 Topic Clustering and Social Network of Backpacker Tourism Content (Sv_yxz7T8rU)

In the video content with ID Sv_yxz7T8rU, 699 records were classified into 10 clusters using both the k-Means and Gaussian Mixture algorithms, with Fast HDBScan also producing 10 clusters using an epsilon value of 0.1. The k-Means and Gaussian Mixture methods efficiently organize the data into clear, well-separated clusters, with k-Means relying on centroid-based grouping and Gaussian Mixture using probabilistic models to capture overlapping cluster structures. Meanwhile, Fast HDBScan's density-based approach allows for more flexible clustering, especially in datasets with noise or varying densities, as the epsilon value controls the threshold for defining clusters. Each algorithm provides a unique perspective on the data, with k-Means and Gaussian Mixture offering more structured classifications, while Fast HDBScan excels in identifying clusters with irregular shapes or densities. This comparative analysis of clustering methods enhances the understanding of viewer engagement patterns and highlights the importance of selecting the most appropriate algorithm based on the dataset's characteristics.

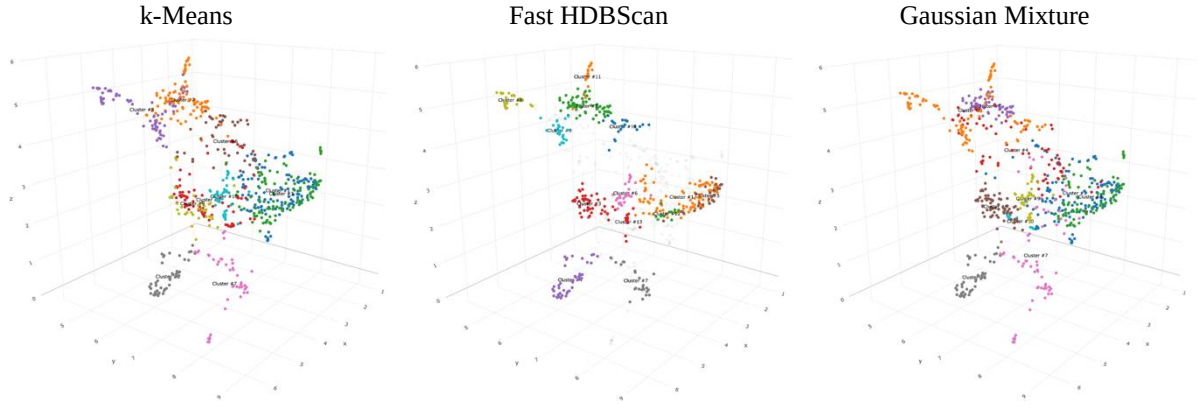

Figure 10. Topic Clustering using k-Means, Fast HDBScan, and Gaussian Mixture (Communalytic)

Figure 10 illustrates the results of topic clustering using three different algorithms: k-Means, Fast HDBScan, and Gaussian Mixture, applied to the same dataset through Communalytic. In the k-Means clustering (first visualization), data points are assigned to centroids based on proximity, resulting in distinct and non-overlapping clusters that emphasize clear group separation. This method works well when the data is evenly distributed. The second visualization, using Fast HDBScan, demonstrates a density-based clustering approach that excels in identifying clusters with varying densities and irregular shapes, making it effective for handling noisy datasets. It highlights the adaptability of this algorithm to different cluster sizes and shapes while managing outliers. The third visualization, based on the Gaussian Mixture model, uses a probabilistic approach, allowing for overlapping clusters where data points can belong to multiple clusters with different probabilities. This flexibility is useful for representing more complex, ambiguous relationships between clusters. By comparing these approaches, it is evident that each algorithm offers distinct advantages, with k-Means providing simplicity and clarity, Fast HDBScan accommodating noise and irregular clusters, and Gaussian Mixture offering a nuanced

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

probabilistic interpretation of the data. These varied methods provide a comprehensive understanding of the dataset's structure, allowing for a tailored analysis depending on the data's characteristics.

In addition, the network consists of 644 actor nodes connected by 51 edges, representing the interactions and relationships within the digital community. The actor nodes symbolize individual users or participants, while the edges capture the connections formed through direct interactions, such as replies or mentions. The relatively small number of edges in comparison to nodes suggests a sparse network with limited engagement between users, where only a few connections are actively contributing to the discussion. This could indicate the presence of a few central figures driving the conversation, while the majority of users remain more passive or isolated in their participation. Analyzing this network structure provides insights into the communication flow, highlighting key actors and identifying potential influencers within the community. Understanding these dynamics is critical for gaining a deeper perspective on how information or discussions propagate and how influential nodes contribute to shaping the overall discourse.

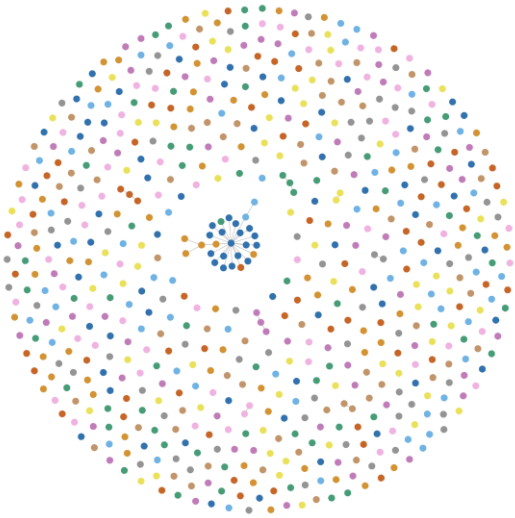

Figure 11. Threaded Discussion Network based on Content id Sv_yxz7T8rU (Communalytic)

Figure 11 illustrates the threaded discussion network based on content ID Sv_yxz7T8rU, as visualized using Communalytic. The network consists of numerous nodes, each representing an individual actor, and edges connecting them to signify interactions such as replies or mentions. The central cluster of densely connected nodes suggests a small group of highly active participants driving the majority of discussions, while the surrounding nodes reflect users who are more passively engaged, with fewer direct interactions. The radiating structure of the network highlights how the conversation expands from the core group outward, with a distinct gap between the central and peripheral actors, indicating that the majority of the network may not be closely involved in the key discussions. This network visualization provides insight into the hierarchical nature of engagement, where a few central figures dominate the discourse, and the majority of users engage sporadically or at the fringes. Analyzing these patterns helps identify the key influencers and the flow of information within the community, offering a deeper understanding of participation dynamics.

The findings from the social network analysis and topic clustering for content ID Sv_yxz7T8rU offer an in-depth view of the interaction patterns and thematic structures within the community. The social network analysis highlights a core group of highly active participants at the center, surrounded by more peripheral users with limited interactions. This suggests that a few key actors play a dominant role in driving discussions, while the majority engage less frequently. The topic clustering, conducted using k-Means, Fast HDBScan, and Gaussian Mixture algorithms, reveals a range of thematic clusters. Each algorithm provides unique insights, with k-Means offering clear, non-overlapping groups, Fast HDBScan identifying irregular, density-based clusters, and Gaussian Mixture capturing overlapping themes. These combined methods reveal both the most discussed topics and how users interact around them, demonstrating that thematic content and user engagement are closely linked. This comprehensive analysis allows for a deeper understanding of how content influences the flow of conversation and which topics resonate most within the digital space.

3.3 Topic Clustering and Social Network of Backpacker Tourism Content (i9t9pbdo-bk)

In the video content with ID i9t9pbdo-bk, 1914 records were classified into 10 clusters using both the k-Means and Gaussian Mixture algorithms, while Fast HDBScan produced 10 clusters with an epsilon value of 0.1. The k-Means algorithm efficiently groups data by assigning records to the nearest centroid, resulting in well-separated, distinct clusters. The Gaussian Mixture model, using probabilistic methods, allows for overlapping clusters, capturing more complex relationships within the data. Meanwhile, Fast HDBScan's density-based approach excels in identifying clusters with varying densities, especially in cases where data points do not conform to clear,

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

spherical distributions. By setting the epsilon to 0.1, Fast HDBScan can flexibly manage noise and anomalies, forming clusters that reflect the natural density variations in the dataset. Comparing these clustering methods provides valuable insights into the structure of viewer engagement, emphasizing the strengths of each algorithm in organizing large and complex datasets, depending on the specific nature of the data being analyzed.

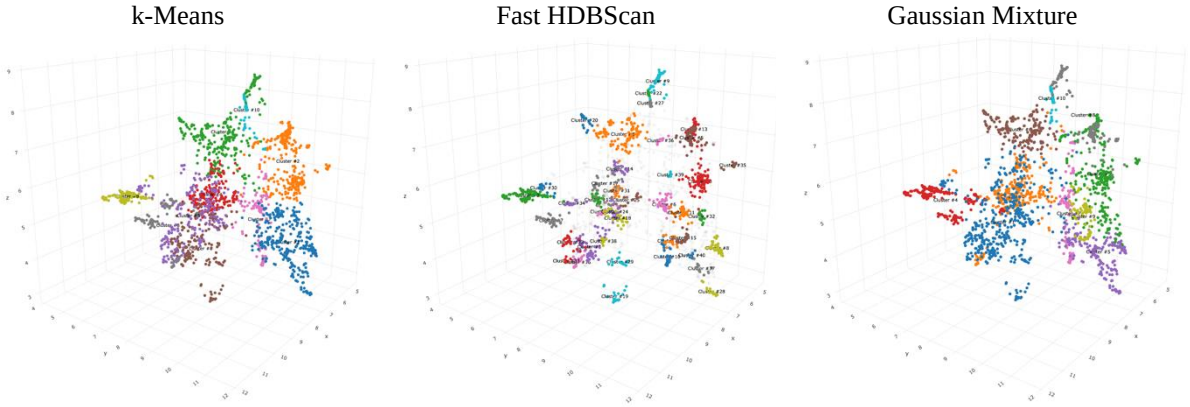

Figure 12. Topic Clustering using k-Means, Fast HDBScan, and Gaussian Mixture (Communalytic)

Figure 12 illustrates the results of topic clustering using three distinct algorithms—k-Means, Fast HDBScan, and Gaussian Mixture—applied to the same dataset via Communalytic. In the first visualization, k-Means efficiently groups data into clearly defined clusters, where data points are assigned to the nearest centroid, resulting in distinct, non-overlapping clusters. This method is effective when the dataset has a clear geometric structure. The second visualization, using Fast HDBScan, demonstrates a density-based clustering approach, which excels in identifying clusters of varying shapes and densities, particularly in the presence of noise and outliers. This method is ideal for datasets with irregular cluster formations and varying densities. The third visualization applies Gaussian Mixture, which allows for probabilistic clustering, enabling overlapping clusters and providing a more flexible model for datasets where points may belong to multiple clusters with varying probabilities. Each of these algorithms offers unique advantages depending on the dataset's complexity, with k-Means providing simplicity, Fast HDBScan handling noisy or complex structures, and Gaussian Mixture offering a nuanced interpretation of overlapping clusters. The comparison of these methods enhances understanding of the underlying data, offering multiple perspectives on clustering depending on the characteristics of the data and analytical goals.

In addition, the network comprises 1,744 actor nodes and 201 edges, representing the intricate web of interactions within the digital community. Each actor node corresponds to an individual participant, while the edges signify the connections formed through their interactions, such as comments, replies, or mentions. The relatively low number of edges in proportion to the nodes suggests a dispersed network, where many users may not engage frequently or are only loosely connected to the core participants. However, the presence of 201 edges indicates active communication among certain users, potentially creating clusters of interaction or influential sub-networks. This configuration reflects the dynamics of user engagement, where a few key actors may serve as central nodes, driving discussions and facilitating the flow of information across the community. Understanding this network structure is essential for identifying patterns of influence, participation, and the spread of information within the digital space, providing deeper insights into the community's behavior and interaction models.

Figure 13. Threaded Discussion Network based on content id i9t9pbdo-bk (Communalytic)

Figure 13 illustrates the threaded discussion network for content ID i9t9pbdo-bk, visualized through Communalytic, showing a complex web of interactions between participants. The network consists of numerous

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

nodes, each representing an individual user, and edges that indicate the connections formed through direct interactions, such as replies or mentions. The central region is densely populated, suggesting a highly active core of users who are closely connected and engage frequently in discussions. As the network extends outward, the density of interactions decreases, with nodes becoming more dispersed, indicating users who are less involved in the central conversation. This structure points to a dynamic interaction model where a small number of highly engaged users influence the majority of discussions, while the wider network consists of more peripheral participants. Such visualizations are valuable for understanding the flow of information, identifying key influencers, and analyzing how engagement is distributed across the community. This detailed analysis helps in revealing the hierarchical nature of the network and how various actors contribute to the overall discussion.

The results from the social network analysis and topic clustering for content ID i9t9pbdo-bk provide valuable insights into the interaction patterns and thematic structures within the discussion. The social network analysis reveals a dense core of highly connected participants, indicating a concentrated group of users who actively shape the conversation, while a broader, less connected group remains on the periphery, engaging less frequently. This structure highlights the role of key actors in driving discussions and influencing information flow. The topic clustering, using k-Means, Fast HDBScan, and Gaussian Mixture algorithms, uncovers distinct thematic clusters that resonate with the community. k-Means offers clear-cut groupings, Fast HDBScan identifies clusters based on density and irregular shapes, and Gaussian Mixture captures overlapping themes, providing a nuanced understanding of the content. Together, these analyses demonstrate the interconnected nature of user interactions and content themes, revealing which topics generate the most engagement and how they spread throughout the network, thus contributing to a comprehensive understanding of the dynamics shaping the online conversation.

4. CONCLUSION

This research presents a novel approach to understanding backpacker tourism by integrating social network analysis with advanced topic clustering techniques, offering fresh insights into digital interactions and thematic trends. Using a structured framework, the study processes 3,575 records from three content IDs (c2ZMFDS_3rU, Sv_yxz7T8rU, and i9t9pbdo-bk), categorizing them into 10 clusters using k-Means, Fast HDBScan, and Gaussian Mixture algorithms. The social network analysis, involving 4,224 actor nodes and 395 edges, reveals the prominence of key users who drive conversations about backpacking experiences, while the broader audience engages less frequently but still contributes to content spread. The novelty lies in applying these computational techniques to backpacker tourism, a domain that has traditionally relied on qualitative methods. This research provides a data-driven perspective on how backpacking topics, such as budget travel, off-the-beaten-path destinations, and sustainable tourism, are discussed and propagated within digital communities. The k-Means algorithm identifies clear-cut thematic clusters, while Fast HDBScan detects more complex, density-based structures, and Gaussian Mixture uncovers overlapping themes, such as the intersection of adventure travel and cultural immersion. By linking topic clustering with social network analysis, the study uncovers how backpacker preferences and discussions evolve in virtual environments, offering actionable insights into which themes resonate most with this travel demographic. This novel approach contributes significantly to backpacker tourism research by offering a detailed, scalable method for understanding traveler behavior, preferences, and engagement within digital spaces, providing valuable implications for content creators and tourism marketers aiming to cater to this niche market.

ACKNOWLEDGMENT

I would like to express my deepest gratitude to the Tourism Study Program, Faculty of Business Administration and Communication, Atma Jaya Catholic University of Indonesia, and Pusat Studi Transformasi Digital dan Pembangunan Pariwisata (PUSDIPAR) for their invaluable support and encouragement throughout the course of this research.

REFERENCES

[1] F. Higgins-Desbiolles, R. A. Scheyvens, and B. Bhatia, “Decolonising tourism and development: from orphanage tourism to community empowerment in Cambodia,” *J. Sustain. Tour.*, vol. 31, no. 12, pp. 2788–2808, 2023, doi: 10.1080/09669582.2022.2039678.

[2] P. Naruetharadhol and N. Gebsombut, “A bibliometric analysis of food tourism studies in Southeast Asia,” *Cogent Bus. Manag.*, vol. 7, no. 1, p. 1733829, Jan. 2020, doi: 10.1080/23311975.2020.1733829.

[3] S. Blasi, S. Fano, S. R. Sedita, and G. Toschi, “A network perspective of cognitive and geographical proximity of sustainable tourism organizations: evidence from Italy,” *Int. J. Contemp. Hosp. Manag.*, vol. 36, no. 2, pp. 478–504, Jan. 2024, doi: 10.1108/IJCHM-03-2022-0366.

[4] E. Surucu-Balci and G. Balci, “Building social capital in cruise travel via social network sites,” *Curr. Issues Tour.*, vol. 26, no. 7, pp. 1096–1111, 2023, doi: 10.1080/13683500.2022.2047904.

[5] M. Valeri and R. Baggio, “Social network analysis: organizational implications in tourism management,” *Int. J. Organ. Anal.*, vol. 29, no. 2, pp. 342–353, Jan. 2021, doi: 10.1108/IJOA-12-2019-1971.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34–48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

[6] F. Espasandin-Bustelo, B. Palacios-Florencio, and J. Sánchez-Rivas García, "CSR intellectual structure in management and tourism," *TQM J.*, vol. 32, no. 3, pp. 521–541, Jan. 2020, doi: 10.1108/TQM-06-2019-0173.
[7] M. De Martino, A. Morvillo, and G. Giordano, "Social network analysis in hospitality and tourism: Guest editorial," *Int. J. Contemp. Hosp. Manag.*, vol. 36, no. 2, pp. 349–357, Jan. 2024, doi: 10.1108/IJCHM-02-2024-161.
[8] R. Baggio, A. Guizzardi, and M. Mariani, "A social network analysis of interlocking directorates in the accommodation sector," *Int. J. Contemp. Hosp. Manag.*, vol. 36, no. 2, pp. 422–437, Jan. 2024, doi: 10.1108/IJCHM-03-2022-0315.
[9] M. Teston *et al.*, "Participatory approaches and Social Network Analysis to analyse the emergence of collective action for rural development: a case study in the Spanish Pyrenees," *Ital. J. Anim. Sci.*, vol. 23, no. 1, pp. 504–522, Dec. 2024, doi: 10.1080/1828051X.2024.2330658.
[10] A. Gour, S. Aggarwal, and M. Erdem, "Reading between the lines: analyzing online reviews by using a multi-method Web-analytics approach," *Int. J. Contemp. Hosp. Manag.*, vol. 33, no. 2, pp. 490–512, Jan. 2021, doi: 10.1108/IJCHM-07-2020-0760.
[11] M. Hocevar and T. Bartol, "Mapping urban tourism issues: analysis of research perspectives through the lens of network visualization," *Int. J. Tour. Cities*, vol. 7, no. 3, pp. 818–844, Jan. 2021, doi: 10.1108/IJTC-05-2020-0110.
[12] Y. Zhang, H. Xu, and H. Yang, "An integrated path framework of tourism and hospitality social entrepreneurship: a systematic literature review," *Int. J. Contemp. Hosp. Manag.*, vol. 36, no. 3, pp. 661–690, Jan. 2024, doi: 10.1108/IJCHM-09-2022-1057.
[13] F. Yüksel and F. Ön, "Sport event research in the hospitality and tourism industry: a bibliometric analysis," *Worldw. Hosp. Tour. Themes*, vol. 16, no. 4, pp. 423–439, Jan. 2024, doi: 10.1108/WHATT-06-2024-0123.
[14] G. Solazzo, G. Elia, and G. Passiante, "Defining the big social data paradigm through a systematic literature review approach," *J. Knowl. Manag.*, vol. 25, no. 7, pp. 1853–1887, Jan. 2021, doi: 10.1108/JKM-10-2020-0801.
[15] A. Arenal, C. Feijoo, A. Moreno, C. Armuña, and S. Ramos, "An academic perspective on the entrepreneurship policy agenda: themes, geographies and evolution," *J. Entrep. Public Policy*, vol. 9, no. 1, pp. 65–93, Jan. 2020, doi: 10.1108/JEPP-06-2019-0056.
[16] H. C. Franz and A. R. Cruz, "Development of a maturity assessment model for sustainable tourism," *Curr. Issues Tour.*, no. May, pp. 1–16, 2024, doi: 10.1080/13683500.2024.2354531.
[17] S. Wu and M. Benson-Rea, "Uncovering the dark side of the sharing economy from a provider's perspective: a bibliometric systematic review," *J. Bus. Ind. Mark.*, vol. 39, no. 6, pp. 1226–1243, Jan. 2024, doi: 10.1108/JBIM-01-2023-0028.
[18] H. Sharma, P. R. Srivastava, S. M. Jasimuddin, Z. J. Zhang, and I. Jebabli, "Privacy concerns in tourism: a systematic literature review using machine learning approach and bibliometric analysis," *Tour. Rev.*, vol. 79, no. 5, pp. 1105–1125, Jan. 2024, doi: 10.1108/TR-10-2022-0517.
[19] P. Kumar, B. Aggarwal, V. Kumar, and H. Saini, "Sustainable tourism progress: a 10-year bibliometric analysis," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2023.2299614.
[20] E. Agyeiwaah and P. Bangwayo-Skeete, "Segmenting and predicting prosocial behaviours among tourists: a latent class approach," *Curr. Issues Tour.*, pp. 1–20, 2023, doi: 10.1080/13683500.2023.2229935.
[21] X. Wang, K. Yang, and T. Liu, "Stock Price Prediction Based on Morphological Similarity Clustering and Hierarchical Temporal Memory," *IEEE Access*, vol. 9, pp. 67241–67248, 2021, doi: 10.1109/ACCESS.2021.3077004.
[22] O. Zhao and J. Wang, "Spatial-temporal behaviour of hikers in the southeastern margin of Qinghai-Tibet Plateau : insights from volunteered geographic information," *Geocarto Int.*, vol. 38, no. 1, p., 2023, doi: 10.1080/10106049.2023.2296183.
[23] Y. Ma and W. Cai, "Social media in ethnographic research: critical reflections on using WeChat in researching Chinese outbound tourists," *Curr. Issues Tour.*, vol. 26, no. 20, pp. 3275–3287, 2023, doi: 10.1080/13683500.2023.2175203.
[24] J. Thieme, M. P. Hampton, C. Stoian, and K. Zigan, "The political economy of backpacker tourism: explorations of tourism actors' embeddedness in Colombia," *Curr. Issues Tour.*, vol. 24, no. 13, pp. 1830–1855, 2021, doi: 10.1080/13683500.2020.1806793.
[25] C. F. Chiang and C. W. Huang, "Online Reviews on Online Travel Agency: Understanding Tourists' Perceived Attributes of Taipei's Economy Hotels," *J. Qual. Assur. Hosp. Tour.*, vol. 23, no. 4, pp. 945–959, 2022, doi: 10.1080/1528008X.2021.1923107.
[26] M. S. Gholamhosseinzadeh, "Theorizing vloggers' approaches and practices in travel vlog production through grounded theory," *J. Hosp. Mark. Manag.*, vol. 32, no. 2, pp. 196–223, 2023, doi: 10.1080/19368623.2023.2164392.
[27] S. Verma, N. Yadav, and R. Chikhalkar, "An integrated measure of eWOM usefulness in the leisure travel: conceptualisation, scale development, and validation," *J. Mark. Commun.*, vol. 29, no. 3, pp. 211–237, 2023, doi: 10.1080/13527266.2021.2004442.
[28] N. Niu, W. Fan, M. Ren, M. Li, and Y. Zhong, "The Role of Social Norms and Personal Costs on Pro-Environmental Behavior: The Mediating Role of Personal Norms," *Psychol. Res. Behav. Manag.*, vol. 16, pp. 2059–2069, 2023, doi: 10.2147/PRBM.S411640.
[29] A. Mandić, S. K. Walia, and S. M. Rasoolimanesh, "Gen Z and the flight shame movement: examining the intersection of emotions, biospheric values, and environmental travel behaviour in an Eastern society," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–23, 2023, doi: 10.1080/09669582.2023.2254950.
[30] 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.
[31] I. Cifci, M. Ogretmenoglu, T. Sengel, T. Demirciftci, and G. Kandemir Altunel, "Effects of Tourists' Street Food Experience and Food Neophobia on Their Post-Travel Behaviors: The Roles of Destination Image and Corona-Phobia," *J. Qual. Assur. Hosp. Tour.*, vol. 00, no. 00, pp. 1–28, 2022, doi: 10.1080/1528008X.2022.2151550.
[32] A. Peters and M. Fuchs, "A relational exploration of tourists' environmental values and their perception of restrictions in protected nature," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–18, 2023, doi: 10.1080/09669582.2023.2295234.
[33] T. B. Luong, "Celebrity involvement, film destination image, place attachment, behavioral intention: the moderating role of e-word of mouth utilitarian function," *Asia Pacific J. Tour. Res.*, vol. 28, no. 9, pp. 949–964, 2023, doi:

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 34−48
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.5969

10.1080/10941665.2023.2283595.
[34] X. Zhang, P. Tavitiyaman, and W. Y. Tsang, "Preferences of Technology Amenities, Satisfaction and Behavioral Intention: The Perspective of Hotel Guests in Hong Kong," *J. Qual. Assur. Hosp. Tour.*, vol. 24, no. 5, pp. 545–575, 2023, doi: 10.1080/1528008X.2022.2070817.
[35] 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.
[36] 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.
[37] F. Huang, C. Yuan, Y. Bi, and J. Lu, "Exploiting Long-Term Dependency for Topic Sentiment Analysis," *IEEE Access*, vol. 8, pp. 221963–221974, 2020, doi: 10.1109/ACCESS.2020.3039963.
[38] P. Madzík, L. Falát, L. Copuš, and M. Valeri, "Digital transformation in tourism: bibliometric literature review based on machine learning approach," *Eur. J. Innov. Manag.*, vol. 26, no. 7, pp. 177–205, 2023, doi: 10.1108/EJIM-09-2022-0531.
[39] J. R. Saura, D. Palacios-Marqués, and D. Ribeiro-Soriano, "Digital marketing in SMEs via data-driven strategies: Reviewing the current state of research," *J. Small Bus. Manag.*, vol. 61, no. 3, pp. 1278–1313, 2023, doi: 10.1080/00472778.2021.1955127.
[40] M. Mariani and M. Borghi, "Environmental discourse in hotel online reviews: a big data analysis," *J. Sustain. Tour.*, vol. 29, no. 5, pp. 829–848, 2020, doi: 10.1080/09669582.2020.1858303.
[41] 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.
[42] C. Kaveski Peres and E. Pacheco Paladini, "Exploring the attributes of hotel service quality in Florianópolis-SC, Brazil: An analysis of tripAdvisor reviews," *Cogent Bus. Manag.*, vol. 8, no. 1, pp. 1–19, 2021, doi: 10.1080/23311975.2021.1926211.