Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 6, December, 2025, Page. 5609-5620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5458

Analyzing User Needs and Recommending Targeted Features for Bi'ih Village Tourism Website Using Text Mining and K-Means Clustering

Mima Artamevia[*1], Muharman Lubis[2], Iqbal Yulizar Mukti[3], Dini Handayani[4]

[1,2,3]Information System, Telkom University, Indonesia
[4]Computer Science, International Islamic University Malaysia, Malaysia

Email: [1]mimaartameviama@student.telkomuniversity.ac.id

Received : Nov 20, 2025; Revised : Nov 23, 2025; Accepted : Nov 23, 2025; Published : Dec 23, 2025

Abstract

Tourism village websites often do not fully reflect user needs, resulting in digital services that cannot be optimally utilized by residents and potential tourists. This situation limits access to information and reduces the effectiveness of tourism promotion efforts, especially in villages that are undergoing digital transformation. This study was conducted to identify the overall needs of users and compile data-based feature recommendations for the development of the Bi'ih Village website as a durian tourism village. The research method used a quantitative approach through the distribution of an online questionnaire to 110 respondents consisting of visitors and residents, with five open-ended questions and several structured questions. The data was analyzed using text mining to find dominant words and themes, as well as the K-Means Clustering technique determined through the Elbow method to group user characteristics. The analysis results showed that there were 2,702 tokens and 677 meaningful words, with the highest demand for government information and visual tourism content. The segmentation process produced three main groups, namely Active Supporters (61.4%), Tech Enthusiasts (27.3%), and Moderate Users (11.4%). This study contributes a data-driven approach to designing more relevant and measurable features for tourism village websites. The impact is expected to increase the adoption of village digital services, strengthen tourism competitiveness, and support the acceleration of the Smart Village concept implementation. The novelty of this study lies in the integration of text mining and clustering as the basis for developing user-oriented feature recommendations.

Keywords : *Bi'ih Village, Clustering, Smart Village, Text Mining, Tourism Website*

1. INTRODUCTION

Digital transformation in villages has become an important part of supporting rural development, not only to meet public service needs, but also to promote local tourism potential as a new source of economic income [1]. Bi'ih Village, located in South Kalimantan and known as the 'durian village', has great potential to attract tourists through effective digital information management. One digital platform that can be a solution is a village website, which serves as an information portal for tourism promotion, government transparency, and community interaction [2]. However, many villages face challenges in determining which website features are appropriate for local needs. Adding irrelevant or excessive features can actually reduce user satisfaction [3]. In Indonesia itself, various studies show that tourism village websites are generally not user-based because feature design still adopts a general approach without considering the social and cultural context and digital preferences of rural communities [4], [5]. Other studies have also found that digital platforms for rural tourism destinations are often ineffective because the features provided do not meet the expectations of tourists or the needs of local businesses [6], [7].

In Bi'ih Village itself, developing a more user-centred village website requires a data-driven approach to understand the specific needs of tourists and residents. Previous research shows that data-

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5458

driven approaches, such as text mining and clustering, can efficiently process survey data to generate relevant recommendations [8]. Text mining is used to extract insights from questionnaire text data, while clustering can group respondents based on their preferences so that user needs patterns can be identified more clearly [9], [10]. Thus, the combination of these two methods provides a great opportunity to generate targeted feature recommendations. A data-driven approach in Smart Villages is increasingly important because text analytics and user behavior segmentation can help system designers understand digital service needs more accurately [11], [12]. In addition, other studies show that digital features and services that do not meet the expectations of users and tourists can significantly hinder the adoption of rural tourism platforms [13], [14].

Text mining is an analytical method that can extract patterns and main topics from questionnaire text data to accurately determine the needs of digital service users [15]. This technique has been widely used in digital tourism destination research to map tourist preferences for web-based information and service features [16]. In the context of village information systems, text mining helps identify feature requests such as harvest calendars, tourist location mapping, and more accessible online administrative services [10]. The results of text mining are then used as the basis for the clustering process. Based on previous research, clustering methods can be used to categorise user characteristics in a system in order to improve the effectiveness of service requirement groupin [17]. Therefore, by combining these two methods, website feature recommendations can be designed based on evidence and be more relevant to the needs of residents and tourists as the main users [18].

Clustering-based approaches are commonly used to group users based on similarities in preferences and characteristics in the use of village digital services [9]. This segmentation helps policymakers determine the most relevant features for each user group [19], For example, tourists who focus on local attractions can be mapped into different clusters from residents who need digital administrative services [20]. Methods such as K-Means have proven effective in distinguishing user behavior profiles in digital tourism development [15]. To ensure that segmentation is neither too few nor excessive, the Elbow method is often applied to determine the optimal number of clusters [21]. However, a hybrid approach combining text mining and clustering to analyze user needs in the context of tourism village websites is still rarely applied, as most previous studies have focused on digital platforms in urban areas, the commercial sector, or e-commerce [12], [22]. This condition indicates a methodological and contextual gap that needs to be bridged through a data-based approach that is more relevant to the social, cultural, and behavioral characteristics of tourism village users.

Previous research based on Smart Villages shows that successful digital transformation in villages requires community participation and a local data-driven approach [23]. In this case, survey data becomes the basis for social legitimacy that strengthens website feature recommendations. In addition, the text mining approach has also been successfully applied in analysing public perception online, demonstrating its effectiveness in supporting sentiment- and user opinion-based digital policies [24]. By combining text mining and clustering, this research not only provides evidence-based policy outcomes, but also supports the development of a more inclusive village information system [25].

This study was designed to address the challenges faced by Bi'ih Village in developing an informative, interactive village website that meets the needs of tourists and the local community. A combination of text mining and clustering methods was used to analyze user preference data in a more systematic and evidence-based manner, so that the resulting feature recommendations are not only general but also reflect the quantitative variations in user characteristics and needs. This approach is expected to increase user adoption, strengthen the transparency of public services, and support more effective village tourism promotion strategies. Thus, the results of this study are not only beneficial for Bi'ih Village but can also serve as a reference for other tourism villages in Indonesia in implementing the concepts of Smart Village and Smart Tourism. Therefore, the main objective of this study is to

Vol. 6, No. 6, December, 2025, Page. 5609-5620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5458

analyze user needs through text mining and the K-Means algorithm to formulate more accurate and data-driven recommendations for the Bi'ih Tourism Village website.

2. METHOD

This study employs a quantitative approach using text mining to extract user needs from survey data and K-Means Clustering to group feature preference patterns for the Desa Bi'ih website. Data were collected through an online questionnaire distributed to tourists interested in durian tourism and residents who use the village's digital services. Text mining processed the open-ended responses to identify dominant word frequencies and feature-related themes, while K-Means Clustering segmented users based on closed-ended ratings [26], [27]. The clustering method was selected due to the diverse and complex characteristics of user data, allowing behaviour-based segmentation to produce more accurate feature recommendations [28]. The research stages included data collection, preprocessing, text mining, Elbow-based cluster determination, K-Means modelling, and feature recommendation development, as shown in Figure 1.

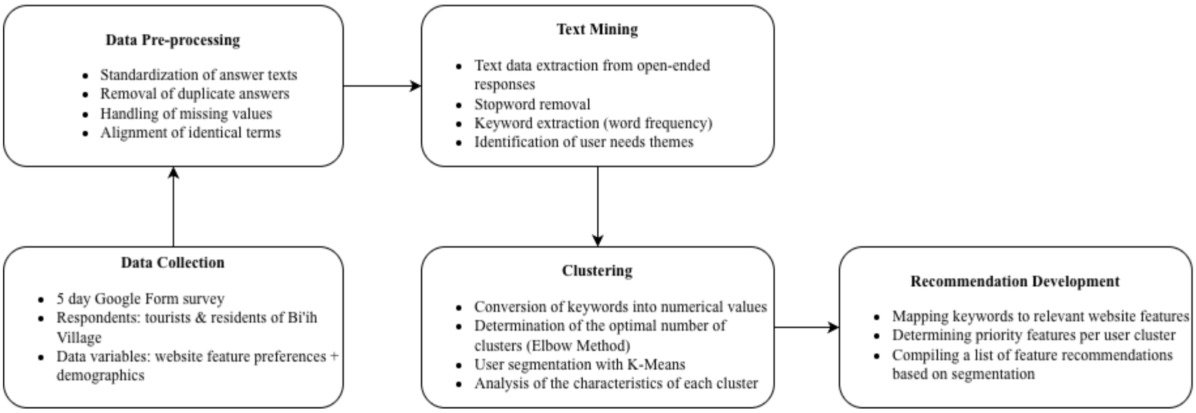

Figure 1. Flow from Survey to Recommendations

2.1. Data Collection

Data collection was carried out through a five-day online survey using Google Forms with accidental sampling to reach users connected to Bi'ih Tourism Village. Respondents consisted of tourists who had visited or were interested in visiting the village and local residents who use its public digital services. The questionnaire gathered demographic data, information-seeking behaviour, and feature preferences. Distribution was conducted via WhatsApp groups and village social media. This digital survey approach was effective for capturing user requirements in a smart tourism context [29]. The collected data were exported to Excel for further analysis, with an example shown in Figure 2.

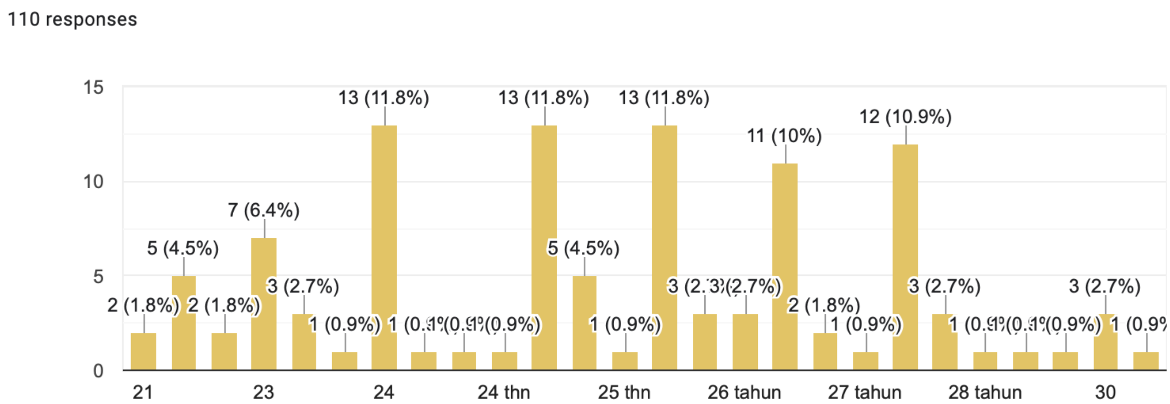

Figure 2. Questionnaire Results

Vol. 6, No. 6, December, 2025, Page. 5609-5620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5458

2.2. Data Pre-processing

Pre-processing is done to ensure data quality and consistency before further analysis, thereby avoiding bias and reducing model prediction error rates [30]. The process begins with text standardization, which involves converting all letters to lowercase, removing non-alphabetic characters, and cleaning up irrelevant punctuation marks. After that, stopword removal is performed using the Indonesian stopword list from NLTK, which provides a set of common words that do not contribute significantly to the analysis. The entire pre-processing sequence is run using the Python programming language with the support of libraries such as NLTK, pandas, and scikit-learn to ensure replication and analytical integrity. This stage is an important foundation in text mining-based research to produce more accurate, structured, and representative analyses [31].

2.3. Text Mining

The text mining stage was conducted to identify the words most frequently used by respondents in describing the development needs of the Bi'ih Village Website. All answers to the five open-ended questions were first combined, then normalised by converting all characters to lowercase and separating words into tokens. All tokens resulting from this processing were entered into a single text corpus, which became the basis for calculating term frequency. Frequency calculations were performed to identify the dominance and trends of topics that emerged in users' perceptions of village digital services. Mathematically, the frequency of a word was calculated using Equation (1) below.

$$Count\ (t) = f\ (K) = \sum_{i=1}^{N} f(t, d_i) \tag{1}$$

Description :

- $Count\ (t)$ = Number of occurrences of the word t in the entire corpus
- $f(t, d_i)$ = Frequency of word occurrence t in document number i
- N = Total number of respondents' documents

Equation (1) is used to calculate the Top 20 words with the highest number of occurrences, so that the most dominant themes can be identified, such as the need for government services, tourist location information, visual content on durian, and other digital service innovations. The results of this analysis will then form the basis for grouping themes and recommending features to be implemented on the Bi'ih Village Website.

2.4. Clustering

In this study, the determination of the optimal number of clusters was carried out using the Elbow Method. This method was first introduced by Thorndike (1953) and is widely used to find the most efficient cluster value K based on the total within-cluster sum of squares (Total WSS). The basic principle of this method is to select a K value that no longer provides a significant decrease in Total WSS, so that the plot graph forms an elbow pattern. The Total WSS calculation is expressed in Equation (2) below [32].

$$Total\ WSS = f\ (K) = \sum_{i=1}^{K} \sum_{E_i \in C_k} (\mu_k - E_i)^2 \tag{2}$$

Description:

- K = Number of clusters tested
- E_i = Element (data) to i in cluster to k
- C_k = Cluster to k

Vol. 6, No. 6, December, 2025, Page. 5609-5620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5458

- μ_k = Centroid (average) of cluster k

The Elbow Method then becomes the basis for determining the best K value before the clustering process is carried out. Once the optimal cluster value has been obtained, the next step is to cluster the data using the K-Means Clustering algorithm. K-Means is the most widely used partitioning clustering method in machine learning to divide data into K groups based on proximity distance [33]. This algorithm aims to minimise variation within each cluster iteratively until the centroid stabilises. The concept of cluster variation measurement (within-cluster variation/WCV) is expressed in Equation (3) below [32]:

$$WCV\ (C_k) = \sum_{E_i \in C_k} (E_i - \mu_k)^2 \tag{3}$$

The total variation of all clusters in K-Means is the sum of the WCV of all clusters, as expressed in Equation (2) [32].

$$Total\ WCV = \sum_{k=1}^{k} \sum_{E_i \in C_k} (E_i - \mu_k)^2 \tag{4}$$

Description:

- μ_k = Cluster centroid
- $WCV\ (C_k)$ = Variation in k clusters
- $Total\ WCV$ = Total variation of all clusters

Thus, the Elbow Method is used as a strategic prerequisite to ensure that the selection of the number of clusters in the K-Means algorithm is not excessive and remains efficient in describing the data structure.

2.5. Recommendation Development

The final stage involves mapping the words with the highest frequency in each cluster to website feature recommendations. For example, if the tourist user cluster most frequently mentions the words ‘location’ and ‘map’, then the priority feature of a durian orchard map will be recommended. Conversely, if the resident cluster mentions ‘letters’ or ‘services’ more often, then the priority feature could be digital village administration services. This approach ensures that feature recommendations are not only based on theory, but also truly reflect the users' voices. The method of developing services based on user needs and data has been proven to support the success of village digital transformation in the context of Smart Villages [18], [34].

3. RESULT

3.1. Data Overview

A total of 110 valid responses were collected through the questionnaire distributed to visitors and potential users of the Desa Bi’ih website. The dataset consists of quantitative ratings related to appearance, usability, and satisfaction, as well as qualitative input obtained from five open-ended questions, producing 550 text entries (110 × 5). These text responses form the basis of the corpus used for the subsequent analysis presented in this section.

3.2. Text Mining Results

Text mining was applied to analyse the qualitative responses provided in the five open-ended items of the questionnaire. These items included suggestions for additional features, development inputs, website management, preferred content, and other user comments. All responses were combined into a

Vol. 6, No. 6, December, 2025, Page. 5609-5620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5458

single corpus for processing. A total of 550 text entries, 2,702 tokens, and 677 meaningful unique words were identified after preprocessing, as summarised in Table 1.

Table 1. Summary of Corpus and Text Variables

Analysis Components	Results
Number of survey questions	38
Open-ended response columns analysed	5
Total text responses (corpus)	550
Total tokens after preprocessing	2702
Meaningful unique words (post-filter)	677

Based on Table 1, it shows that the size of the data collected is representative enough to describe the patterns of website user needs. The five columns of open-ended answers were retained because they all contained direct input regarding the features and content that users expected from the website. Next, the frequency of words from the open-ended responses (fill-in-the-blank) was calculated to identify the most frequently occurring words. The results of the analysis show that the word "desa" is the word with the highest frequency (184x). The words "video", "fitur", and "tambahkan" also have a high frequency, which indicates a need for additional features and stronger visual content on the website. In addition, words such as "pemerintah", "warga" and "informasi" appeared, indicating that the website is perceived as an official channel for village public services.

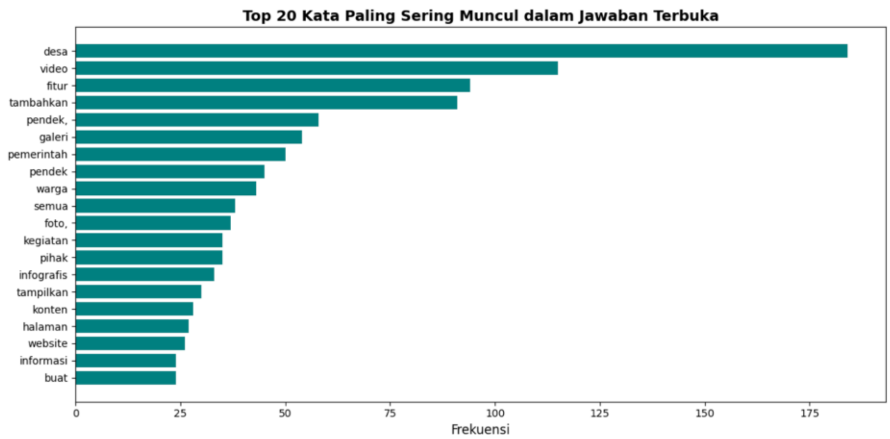

Figure 3. Top 20 Most Frequently Occurring Words

Figure 3 presents the Top 20 most frequently occurring words extracted from the 550 text entries in the corpus. The word "desa" appears with the highest frequency at 184 occurrences, followed by "video" (115), "galeri" (54), "infografis" (33), "fitur" (94), and "tambahkan" (91). Other frequently occurring terms include "pemerintah" (50), "warga" (43), "konten" (28), and "informasi" (24). These frequency counts represent the distribution of dominant terms found in the open-ended responses and serve as the basis for further thematic grouping presented in Table 2.

Table 2 Grouping of Themes Based on Word Frequency

Themes	Frequency
Government	315
Visual Content	225
Digital Features	174
Information	117
Public Services	40
SMEs & Economy	32

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 6, December, 2025, Page. 5609-5620
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5458

Table 2 shows the distribution of word frequencies grouped into six themes: Governance (315 occurrences), Visual Content (225), Digital Features (174), Information (117), Public Services (40), and SMEs & Economy (32). These frequency values represent the thematic classification derived from the text corpus and form the basis for further interpretation in the discussion section.

3.3. Clustering Analysis

Clustering analysis was performed to group respondents based on similarities in their ratings of website appearance, ease of access, loading speed, additional features, and digital services. All variables were processed in numeric form and normalized to ensure consistency before clustering. The next step involved determining the optimal number of clusters using the Elbow Method, as shown in Figure 4.

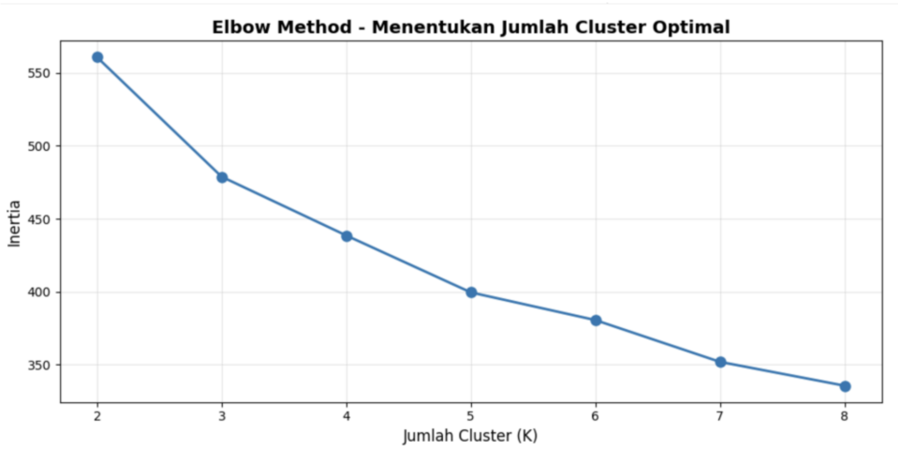

Figure 4. Cluster Determination Using Elbow

The first step in determining the optimal number of clusters is done using the Elbow method, which involves calculating the Within-Cluster Sum of Squares (WCSS) for various values of k. Based on Figure 3, which is the Elbow graph, it shows that the elbow point is at k=3. This means that increasing the number of clusters above 3 does not cause a significant decrease in WCSS, so k=3 is chosen as the efficient and interpretative number of clusters with simple segmentation but covering high variability, which is more applicable in the context of public services. After setting k=3, the K-Means clustering algorithm is applied to divide respondents into three different segments.

3.4. Cluster Profiling

The typeface used in the script is Times New Roman, except for the writing of e-mail accounts. Other fonts can be used if needed for certain purposes. After the clustering process using the K-Means method, each cluster was then given a behavior label to facilitate interpretation of user needs. Labeling was done by observing the pattern of response tendencies in each cluster. The results are as follows.

- Cluster 1 = Active Supporter (Agree Tourism Features)
- Cluster 2 = Tech Enthusiast (Strongly Agree All)
- Cluster 3 = Moderate User (Neutral/Undecided)

The distribution of respondents in each cluster is shown in Table 3 as follows.

Table 3 Profile of Requirements for Each Cluster

Cluster Name	Persentase
Active Supporter (Agree with Tourism Features)	61,4%
Tech Enthusiast (Strongly Agree with All)	27,3%
Moderate User (Neutral/Undecided)	11,4%

Vol. 6, No. 6, December, 2025, Page. 5609-5620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5458

Based on Table 3, it can be seen that the majority of respondents (61.4%) fall into the Active Supporter category, which is a group that strongly supports tourism and digital MSME features on websites. A total of 27.3% of respondents fall into the Tech Enthusiast category, which are digital users who are enthusiastic about all aspects of modern feature development. Meanwhile, 11.4% of respondents are Moderate Users, which are groups that are still neutral or hesitant in utilizing digital features.

3.5. Integrated Feature Mapping

To integrate the findings from text mining and clustering, a feature mapping process was conducted to align the most frequently mentioned user needs with the characteristics of each user segment. This integration provides a structured overview of priority areas for website development and identifies which features are most relevant to the dominant patterns emerging from the data. The results of this mapping are presented in Table 4.

Table 4 Mapping the Development of Features for the Bi'ih Village Website

Development Priorities	Data	User Needs	Proposed Features
Tourism Visualization and Local Potential	The words "video," "foto," and "galeri" appear 225 times.	More attractive appearance for tourists and residents	Durian photo/video gallery Tourism profile page Village event highlights
Strengthening Digital Public Services	The theme of government appears 315 times.	Website as an official information channel	Online mail service Village administration contact center Village structure and official announcements
Digital Economy & MSME Promotion	The word "MSME" appears 23 times.	Expectations for promoting local products	MSME product catalog Online ordering + QRIS/e-wallet
Simple Navigation & User Education	Frequent terms: "mudah," "alur"	Moderate Users (11.4%) need assistance with usage	User-Friendly Interface Interactive guide to writing letters Quick access FAQ
Interaction-Based Tourism Features	Active/Tech cluster support	More efficient access to tourism	Interactive map of durian locations Tourist review & rating system
Sustainable Digital Management	Consistent requests for feature improvements	Content availability must be stable	Formation of Village Digital Teams / Digital Village-Owned Enterprises

Table 4 presents the distribution of feature priorities generated from the combined analysis, showing how frequently mentioned needs correspond to specific development components on the website. The mapping illustrates the relationship between the extracted keywords, user expectations, and the set of features that align directly with the observed patterns in the dataset.

4. DISCUSSIONS

The results of word frequency analysis and theme grouping show that governance is the most dominant user requirement, indicating high expectations for transparency and official information, in line with various studies on digital governance in rural areas [35], [36]. A similar pattern was also found in research on village portals in Thailand and Vietnam, where government content was also the focus of users' attention [37], [38]. The theme of Visual Content, which emerged as the second greatest need,

Vol. 6, No. 6, December, 2025, Page. 5609-5620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5458

supports the findings in digital tourism research that visual elements have a strong influence on tourist engagemen [39]. Findings regarding digital feature requirements indicate that users are increasingly demanding interactive online services.

Then, the clustering analysis produced three groups of users with different levels of digital readiness. This pattern shows that rural users are not homogeneous in their openness to new technologies. This is in line with previous studies showing that user segmentation on tourism platforms and digital public services is often influenced by digital literacy and technology usage habits [40]. These differences in preferences are also evident in the context of rural tourism in other countries, where more tech-savvy user groups tend to want advanced features, while moderate groups prefer a simpler interface [41].

The integration of text mining and clustering methods provides a clearer and more consistent picture of user needs. The two analysis results support each other, allowing feature priorities to be determined in a more focused manner. These findings are in line with research showing that the application of AI techniques in text analysis can improve the accuracy of identifying user needs in digital tourism systems [42]. Thus, the contribution of AI in this study strengthens the development of inclusive tourism villages through more precise mapping of the needs of user groups with different levels of technological readiness.

The resulting feature mapping provides a more systematic direction for development based on the real needs of users. The priority on digital governance supports transparency in public services and community participation, as demonstrated by empirical evidence of the role of digital villages in improving the delivery of public services in rural areas [43]. Strengthening visual content is relevant to findings that state that the quality of digital marketing contributes to the image of rural destinations and tourists' intention to visit [44]. In addition, support for MSME features and simple navigation needs is consistent with studies on MSME digital readiness and transformation that emphasize the importance of technological readiness and local digital strategies to improve the performance of micro and small businesses [45].

5. CONCLUSION

This study successfully identified user needs for the development of the Bi'ih Village website through a text mining and clustering approach. The results of the analysis show that the village website is seen not only as a channel for government information, but also as the main medium for promoting durian tourism and supporting local economic activities. The greatest user needs relate to transparency in governance, the presentation of attractive visual content, and more interactive digital service features. In addition, segmentation analysis identified three user groups with different levels of digital readiness, namely the highly tech-savvy group, the active supporters group, and the moderate users group who need simpler navigation. These findings confirm that feature development cannot be carried out uniformly, but must take into account the digital profile of each segment.

This study makes an important contribution by revealing user needs more accurately through a hybrid method of text mining and clustering, and strengthening the development of data-based rural informatics relevant to tourist villages. This approach not only produces more targeted feature recommendations, but also shows how AI-based analytical techniques can help to understand the digital behavior of rural communities in greater depth. Further research can directly implement the designed features and assess their effectiveness through usability testing, as well as conduct trials in multiple villages to ensure the results can be generalized.

Vol. 6, No. 6, December, 2025, Page. 5609-5620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5458

ACKNOWLEDGEMENT

The researchers would like to express their sincere gratitude to all parties who have provided support and contributions in the implementation of this research. The various forms of guidance, input, and valuable experiences received have played an important role in perfecting this research. Deep gratitude is also expressed for the encouragement and involvement of all parties throughout the research process.

REFERENCES

[1] V. Emerllahu and D. Bogataj, "Smart Villages as infrastructure of rural areas: Literature review and research agenda," *IFAC-PapersOnLine*, vol. 58, no. 3, pp. 268–273, Jun. 2024, doi: 10.1016/j.ifacol.2024.07.162.
[2] J. Martinez-Gil *et al.*, "Framework for Data Analysis in the Context of the Smart Villages," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, pp. 1–5. doi: 10.1007/978-3-030-58957-8_3.
[3] A. Paniagua, "Smart and Novelty Villages as the Quality Place of Virtuality," *Sustainability (Switzerland)*, vol. 15, no. 15, Aug. 2023, doi: 10.3390/su151511702.
[4] Amanda Fitria Desvita, Ariska Nurul Habibah, Dwi Yuliana, and Ilham Albana, "Analisa dan Perancangan Website Desa Sebagai Sistem Informasi dan Inovasi di Desa Sokawera, Kecamatan Somagede," *Merkurius : Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 4, pp. 39–51, Jun. 2024, doi: 10.61132/merkurius.v2i4.135.
[5] W. Abbas and S. Sutrisno, "Pengembangan Website Desa sebagai Sistem Informasi dan Inovasi di Desa Indu Makkombong, Kabupaten Polewali Mandar," *Jurnal Abdi Masyarakat Indonesia*, vol. 2, no. 2, pp. 505–512, Feb. 2022, doi: 10.54082/jamsi.276.
[6] X. Liu, M.-Y. Li, Y.-M. Ma, T.-H. Gao, and D.-N. Yuan, "Personalized tourism product design focused on tourist expectations and online reviews: An integrated MCDM method," *Comput Ind Eng*, vol. 188, p. 109860, 2024, doi: 10.1016/j.cie.2023.109860.
[7] H. Wang and J. Yan, "Effects of social media tourism information quality on destination travel intention: Mediation effect of self-congruity and trust," *Front Psychol*, vol. 13, pp. 1–14, Dec. 2022, doi: 10.3389/fpsyg.2022.1049149.
[8] S. Tiruveedhula, C. M. Rani, and V. Narayana, "A Survey on Clustering Techniques for Big Data Mining," *Indian J Sci Technol*, vol. 9, pp. 1–12, Feb. 2016, doi: 10.17485/ijst/2016/v9i3/75971.
[9] T.-L. Wang and M. Azizurrohman, "Segmenting International Tourists in Indonesia: A Cluster Analysis of Preferences, Motivations, and Behaviors," *Journal of Economics Research and Social Sciences*, vol. 9, pp. 93–107, Feb. 2025, doi: 10.18196/jerss.v9i1.23582.
[10] T. Apriliana, N. Aulia Maknunah, H. Muhammad Nizam, R. Mardhatillah, and R. Fatwa Nova, "Analyzing Tourist Satisfaction Using Factor Analysis and Text Mining: An Ecotourism Study in Girpasang Village," *Enthusiastic International Journal Of Applied Statistics And Data Science*, vol. 4, no. 1, pp. 37–50, Apr. 2024, doi: 10.20885/enthusiastic.vol4.iss1.art4.
[11] E. A. Muhtar, A. Abdillah, I. Widianingsih, and Q. M. Adikancana, "Smart villages, rural development and community vulnerability in Indonesia: A bibliometric analysis," *Cogent Soc Sci*, vol. 9, no. 1, pp. 1–25, 2023, doi: 10.1080/23311886.2023.2219118.
[12] W. Wu, C. Xu, M. Zhao, X. Li, and R. Law, "Digital Tourism and Smart Development: State-of-the-Art Review," *Sustainability (Switzerland)*, vol. 16, no. 23, pp. 1–19, Dec. 2024, doi: 10.3390/su162310382.
[13] I. Sustacha, J. Baños-Pino, and E. Del Valle, "The role of technology in enhancing the tourism experience in smart destinations: A meta-analysis," *Journal of Destination Marketing & Management*, vol. 30, Dec. 2023, doi: 10.1016/j.jdmm.2023.100817.
[14] Y. Zhang, M. Sotiriadis, and S. Shen, "Investigating the Impact of Smart Tourism Technologies on Tourists' Experiences," *Sustainability (Switzerland)*, vol. 14, no. 5, Mar. 2022, doi: 10.3390/su14053048.

Vol. 6, No. 6, December, 2025, Page. 5609-5620
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5458

[15] I. C. Chang, J. S. Horng, C. H. Liu, S. F. Chou, and T. Y. Yu, "Exploration of Topic Classification in the Tourism Field with Text Mining Technology—A Case Study of the Academic Journal Papers," *Sustainability (Switzerland)*, vol. 14, no. 7, Apr. 2022, doi: 10.3390/su14074053.
[16] Q. Li, S. Li, S. Zhang, J. Hu, and J. Hu, "A review of text corpus-based tourism big data mining," Aug. 01, 2019, *MDPI AG*. doi: 10.3390/app9163300.
[17] M. Lubis, R. Fauzi, and M. Azani Hasibuan, "Categorization for the Security Awareness Domain and Resource (SADAR) Model in the Organization: A Clustering Statistical Analysis," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jul. 2019, pp. 1–7. doi: 10.1088/1742-6596/1235/1/012041.
[18] A. P. E. Susilowati, R. Rachmawati, and R. Rijanta, "Smart village concept in Indonesia: ICT as determining factor," *Heliyon*, vol. 11, no. 1, Jan. 2025, doi: 10.1016/j.heliyon.2025.e41657.
[19] Ł. Satoła and A. Milewska, "The Concept of a Smart Village as an Innovative Way of Implementing Public Tasks in the Era of Instability on the Energy Market—Examples from Poland," Jul. 01, 2022, *MDPI*. doi: 10.3390/en15145175.
[20] S.-S. Kim, W. Shin, and H.-W. Kim, "Unravelling long-stay tourist experiences and satisfaction: text mining and deep learning approaches," *Current Issues in Tourism*, vol. 28, no. 3, pp. 492–510, Feb. 2025, doi: 10.1080/13683500.2024.2327840.
[21] M. Deldadehasl, H. H. Karahroodi, and P. Haddadian Nekah, "Customer Clustering and Marketing Optimization in Hospitality: A Hybrid Data Mining and Decision-Making Approach from an Emerging Economy," *Tourism and Hospitality*, vol. 6, no. 2, Jun. 2025, doi: 10.3390/tourhosp6020080.
[22] S. I. U. Arul and I. Sanjaya, "Penerapan Algoritma K-Means Clustering pada Sentimen Pengunjung Desa Wisata Hanjeli," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 5, no. 1, pp. 126–133, May 2024, doi: 10.37859/coscitech.v5i1.6955.
[23] A. Junaidi, M. S. Bin Othman, S. Z. Mohd Hashim, M. M. Mohamad, D. Danial Kesa, and A. Nurfikri, "Smart villages: a systematic review of trends, models, and metrics," *Cogent Soc Sci*, vol. 11, no. 1, 2025, doi: 10.1080/23311886.2025.2492833.
[24] F. S. Lubis, M. Lubis, L. Hakim, and H. Fakhrurroja, "The Text Mining Analysis Approach for Electronic Information and Transaction (ITE) Implementation Based on Sentiment in the Social Media," in *Lecture Notes in Networks and Systems*, Springer Science and Business Media Deutschland GmbH, 2023, pp. 263–271. doi: 10.1007/978-981-19-7660-5_23.
[25] W. Serrano, "Intelligent recommender system for big data applications based on the random neural network," *Big Data and Cognitive Computing*, vol. 3, no. 1, pp. 1–29, Mar. 2019, doi: 10.3390/bdcc3010015.
[26] S. Li, F. Liu, Y. Zhang, B. Zhu, H. Zhu, and Z. Yu, "Text Mining of User-Generated Content (UGC) for Business Applications in E-Commerce: A Systematic Review," Oct. 01, 2022, *MDPI*. doi: 10.3390/math10193554.
[27] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability (Switzerland)*, vol. 14, no. 12, Jun. 2022, doi: 10.3390/su14127243.
[28] A. Rimakka and R. Aras, "User Segmentation Based on Purchasing Habits and Preferences on the Amazon Platform Using K-Means Clustering," *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, vol. 13, pp. 103–109, Dec. 2023, doi: 10.35585/inspir.v13i2.63.
[29] U. Gretzel, M. Sigala, Z. Xiang, and C. Koo, "Smart tourism: foundations and developments," *Electronic Markets*, vol. 25, no. 3, pp. 179–188, Sep. 2015, doi: 10.1007/s12525-015-0196-8.
[30] M. Hosseinzadeh *et al.*, "Data cleansing mechanisms and approaches for big data analytics: a systematic study," *J Ambient Intell Humaniz Comput*, vol. 14, pp. 1–13, Nov. 2021, doi: 10.1007/s12652-021-03590-2.
[31] M. Lamba and M. Margam, "Text Pre-Processing," in *Text Mining for Information Professionals*, 2022, ch. Chapter 3, pp. 79–103. doi: 10.1007/978-3-030-85085-2_3.
[32] S. Renjith, A. Sreekumar, and M. Jathavedan, "An empirical research and comparative analysis of clustering performance for processing categorical and numerical data extracts from social media," *Acta Scientiarum - Technology*, vol. 44, Jan. 2022, doi: 10.4025/ACTASCITECHNOL.V44I1.58653.

[33] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *J R Stat Soc Ser C Appl Stat*, vol. 28, no. 1, pp. 100–108, 1979, doi: 10.2307/2346830.

[34] E. A. Manapa Sampetoding and M. Er, "Digital Transformation of Smart Village: A Systematic Literature Review," *Procedia Comput Sci*, vol. 239, no. 2023, pp. 1336–1343, 2024, doi: 10.1016/j.procs.2024.06.304.

[35] Suprapti, R. Ardhian Nugroho, and S. Galan Prakoso, "Contributing Factors of Smart Village Implementation in Rural Area: A Systematic Literature Review," *Publik: Jurnal Ilmu Administrasi*, vol. 14, no. 1, pp. 77–95, Jun. 2025, doi: 10.31314/pjia.14.1.77-95.2025.

[36] F. Liao, Y. Zheng, and D. Yang, "Digital transformation in rural governance: unraveling the micro-mechanisms and the role of government subsidies," *Humanit Soc Sci Commun*, vol. 12, no. 1, Dec. 2025, doi: 10.1057/s41599-025-05716-9.

[37] D. Sagarik, P. Chansukree, W. Cho, and E. Berman, "E-government 4.0 in Thailand: The role of central agencies," IOS Press, 2018. doi: 10.3233/IP-180006.

[38] T. Thi Uyen Nguyen, P. Van Nguyen, H. Thi Ngoc Huynh, G. Q. Truong, and L. Do, "Unlocking e-government adoption: Exploring the role of perceived usefulness, ease of use, trust, and social media engagement in Vietnam," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 2, p. 100291, 2024, doi: https://doi.org/10.1016/j.joitmc.2024.100291.

[39] M. A. Widianti and R. Sepriandana, "Leveraging new media for the branding of pinge tourism village," *Priviet Social Sciences Journal*, vol. 5, no. 10, pp. 81–90, Oct. 2025, doi: 10.55942/pssj.v5i10.656.

[40] I. Mutmainah *et al.*, "Analysis of Factors Influencing Digital Transformation of Tourism Villages: Evidence from Bogor, Indonesia," *Tourism and Hospitality*, vol. 6, no. 2, p. 23, Jun. 2025, doi: 10.3390/tourhosp6020057.

[41] I. G. Saputra, "Digitalization in Rural Tourism Products: From Stages to Managerial Readiness," *International Journal of Human Research and Social Science Studies*, vol. 01, Dec. 2024, doi: 10.55677/ijhrsss/01-2024-Vol01I6.

[42] T. Mou and H. Wang, "Online comments of tourist attractions combining artificial intelligence text mining model and attention mechanism," *Sci Rep*, vol. 15, no. 1, p. 1121, 2025, doi: 10.1038/s41598-025-85139-3.

[43] Z. Mao, X. Zhu, Q. Zou, and W. Jin, "How Can Digital Villages Improve Basic Public Services Delivery in Rural Areas? Evidence from 1840 Counties in China," *Agriculture (Switzerland)*, vol. 14, no. 10, Oct. 2024, doi: 10.3390/agriculture14101802.

[44] S. Rodrigues, R. Correia, R. Gonçalves, F. Branco, and J. Martins, "Digital Marketing's Impact on Rural Destinations' Image, Intention to Visit, and Destination Sustainability," *Sustainability (Switzerland)*, vol. 15, no. 3, pp. 1–20, Feb. 2023, doi: 10.3390/su15032683.

[45] L. Anatan and Nur, "Micro, Small, and Medium Enterprises' Readiness for Digital Transformation in Indonesia," *economies*, vol. 11, pp. 1–21, May 2023, doi: 10.3390/economies11060156.