SEMAR JOURNAL
(Scientific Education in Machine Learning and Artificial Intelligence for Research

Development and Evaluation of a Content-Based Tourism Recommendation System Using TF-IDF and Cosine Similarity: A Case Study of Labuan Bajo, Indonesia

Laizza Natta Fatdaja[1]✉, Hafizurrahman[2], Uslan[3], Izzan Khairul bin Abdullah Faris[4], Alesha Sofea binti Iskandar Hakim[5]

[1,2]Faculty of Communication and Informatics, Universitas Muhammadiyah Surakarta, Indonesia
[3]Faculty of Teacher Training and Education, Universitas Muhammadiyah Kupang, Indonesia
[4,5]Malaysian Institute of Information Technology, Universiti Kuala Lumpur, Malaysia

doi: xxxx
Received: April 12th, 2025 | Revised: May 10th, 2025 | Accepted: May 18th, 2025
Available Online: May 23rd, 2025 | Published Regularly: May, 2025

Abstract

The increasing number of tourism destinations in Labuan Bajo provides tourists with diverse travel alternatives but also creates challenges in identifying destinations that match their individual preferences. This study aims to develop a natural tourism recommendation system using an item-based Content-Based Filtering approach integrated with Term Frequency–Inverse Document Frequency (TF-IDF) and Cosine Similarity. The proposed system utilizes a dataset comprising 50 natural tourism destinations, with textual descriptions serving as the primary features for recommendation generation. The research involves data preprocessing, TF-IDF vectorization, similarity calculation, recommendation generation, and performance evaluation using Precision@5. The experimental results demonstrate that TF-IDF effectively represents distinctive destination characteristics, while Cosine Similarity identifies relationships between destinations based on their textual features. The highest similarity score of 0.91 was obtained between Komodo Island and Rinca Island, indicating substantial similarities in their descriptive characteristics. Furthermore, the recommendation system achieved an average Precision@5 of 0.80 across five evaluation scenarios, with the highest precision of 1.00 for Rinca Island and the lowest of 0.60 for Pink Beach. The proposed system was also implemented as an interactive web application that enables users to explore ranked destination recommendations based on their selected reference destinations. These findings demonstrate the potential of integrating TF-IDF and Cosine Similarity to support personalized natural tourism recommendations and facilitate destination exploration in Labuan Bajo.

Keywords: Content-Based Filtering; Cosine Similarity; TF-IDF; Tourism Recommendation System; Labuan Bajo.

This is an open access article under the CC-BY license.

✉**Corresponding Author:**
Laizza Natta Fatdaja, Faculty of Communication and Informatics, Universitas Muhammadiyah Surakarta, Indonesia
Email: l300240049@student.ums.ac.id

Introduction

Labuan Bajo, located in West Manggarai Regency, East Nusa Tenggara, Indonesia, has been designated as one of the country's Super Priority Tourism Destinations (Destinasi Pariwisata Super Prioritas/DPSP) (Islahuddin et al., 2022; Pahlevi, 2024). Its growing recognition as an international tourism destination is reflected in the increasing number of domestic and international visitors. According to tourism statistics in 2023, the Labuan Bajo and West Manggarai region recorded more than 423,000 tourist arrivals (Ardin, 2024a). Furthermore, Komodo National Park, which serves as one of the

region's primary tourism attractions, received approximately 300,488 visitors throughout the same year (Ardin, 2024b). These figures demonstrate the substantial potential of Labuan Bajo's natural tourism sector and its growing importance in supporting regional economic development (Suharyanto et al., 2024; Teguh & Scott, 2026). The region offers a diverse range of natural attractions, including Komodo National Park, Padar Island, Pink Beach, Rinca Island, and various marine tourism destinations. Each destination possesses distinctive geographical characteristics, ecological features, and recreational activities that provide different experiences for tourists. However, the increasing availability of tourism alternatives also introduces challenges for visitors in identifying destinations that correspond to their individual interests, expectations, and preferences. Tourists may experience difficulties in comparing destinations based on their characteristics, particularly when relevant information is distributed across multiple sources (Al Fararni et al., 2021; Roetzel, 2019). Consequently, an intelligent and personalized recommendation system is necessary to facilitate destination selection by providing relevant alternatives based on the characteristics of previously selected tourism destinations (Hamid et al., 2021; Renjith et al., 2020).

Recommendation systems have become an important information technology solution for addressing information overload and supporting personalized decision-making in various application domains, including tourism (Alyari & Jafari Navimipour, 2018; Roy & Dutta, 2022). In general, recommendation systems utilize available information about users, items, or their interactions to identify and present relevant alternatives (Ko et al., 2022; Zhang et al., 2019). Among the established recommendation techniques, Content-Based Filtering (CBF) determines recommendations by analyzing the characteristics and descriptive attributes of items and identifying similarities between them (Javed et al., 2021; Lops et al., 2019). This approach is particularly relevant to tourism applications because destinations are commonly represented through textual descriptions containing information about geographical conditions, natural attractions, recreational activities, and distinctive environmental characteristics (Al Fararni et al., 2021; Huda et al., 2021). By analyzing these textual attributes, CBF can identify destinations with similar characteristics and generate recommendations without depending entirely on historical interactions among multiple users (Panda & Ray, 2022; Yuan & Hernandez, 2023). Previous research has demonstrated the applicability of this approach in tourism recommendation systems (Oyadila et al., 2025; Salsabilla & Utomo, 2025). Pratiwi et al. (2025) developed a recommendation system for tourist attractions in Aceh Tamiang using Content-Based Filtering by integrating Term Frequency-Inverse Document Frequency (TF-IDF) weighting and Cosine Similarity to measure the similarity of textual destination information. Similarly, Pasaribu and Sitompul (2023) investigated tourism recommendation systems using a user-based Collaborative Filtering approach to identify similarities in tourist behavior and preferences. These studies demonstrate the potential of recommendation algorithms to improve the relevance of tourism information and support destination selection (Faurina & Sitanggang, 2023; Tarigan et al., 2023). In the Labuan Bajo context specifically, earlier recommendation studies relied on k-

nearest neighbor classification of discrete destination attributes (Panjaitan et al., 2022), whereas more recent work has combined TF-IDF with Sentence-BERT representations (Kusumaningrum & Rohman, 2026) and compared similarity measures for culinary tourism recommendations (Masmur et al., 2026). Nevertheless, differences in recommendation strategies, data characteristics, and similarity measurement approaches indicate the need for further investigation into content-oriented recommendation models, particularly in environments where destination descriptions constitute the primary source of information (Albab & Rohman, 2026; Huda et al., 2024).

One of the fundamental challenges in implementing Content-Based Filtering is transforming unstructured textual descriptions into numerical representations that can be systematically compared (Kadhim, 2019; Kowsari et al., 2019). The quality of this representation directly influences the system's ability to distinguish relevant characteristics among tourism destinations. In this context, TF-IDF provides an effective weighting mechanism by considering the frequency of a term within a document and its relative importance across the entire document collection (Kim & Gil, 2019; Qaiser & Ali, 2018). Consequently, terms that are distinctive to particular destinations receive greater weights, allowing the system to identify distinguishing characteristics more effectively (Chen, 2024; Qaiser & Ali, 2018). The resulting numerical representations can subsequently be compared using Cosine Similarity, a vector-based similarity measurement that evaluates the angular relationship between two vectors. Unlike distance-based measurements that may be influenced by vector magnitude, Cosine Similarity focuses on the orientation of vectors, making it suitable for comparing textual descriptions with different lengths (Wang & Dong, 2020). The integration of TF-IDF and Cosine Similarity therefore provides a computational framework for identifying tourism destinations with comparable descriptive characteristics and generating recommendations based on their content similarity.

Despite the established application of these techniques, an important research issue remains concerning how textual feature representation and similarity measurement can be integrated to produce relevant recommendations for natural tourism destinations with diverse characteristics. In particular, destinations that share broad geographical categories may possess substantially different attractions and recreational activities. For example, conservation-oriented islands and islands characterized by panoramic landscapes may belong to the same general category but offer fundamentally different tourism experiences. A recommendation mechanism based solely on general category classification may therefore generate alternatives that do not adequately reflect the distinctive characteristics of the selected destination. Addressing this issue requires a recommendation approach capable of identifying similarities at the descriptive feature level rather than relying exclusively on broad destination classifications. Furthermore, the relevance of the resulting recommendations needs to be systematically evaluated to determine whether the identified similarities correspond to tourists' preferences (Silveira et al., 2019; Zangerle & Bauer, 2022).

Based on these considerations, this study develops a natural tourism recommendation

system for Labuan Bajo using an item-based Content-Based Filtering approach integrated with TF-IDF and Cosine Similarity. The proposed system processes textual descriptions of tourism destinations through several preprocessing stages, including duplicate elimination, case folding, stopword removal, word normalization, and stemming (HaCohen-Kerner et al., 2020). The cleaned textual data are subsequently transformed into numerical feature vectors using TF-IDF, while Cosine Similarity is employed to calculate the degree of similarity between destinations. Recommendations are then generated by ranking destinations according to their similarity scores. To assess recommendation relevance, the system employs the Precision@K evaluation metric by comparing the generated recommendations with user rating data (Herlawati et al., 2026; Silveira et al., 2019). The main contribution of this research is the implementation and evaluation of an integrated text-based recommendation framework that identifies similarities among natural tourism destinations based on their descriptive characteristics. Through this approach, the study aims to provide tourists with more relevant and personalized destination alternatives, facilitate informed tourism decisions, and support the exploration of the diverse natural tourism potential of Labuan Bajo.

Method

This study employs an item-based Content-Based Filtering (CBF) approach to develop a natural tourism recommendation system for Labuan Bajo, Indonesia. The method integrates Term Frequency–Inverse Document Frequency (TF-IDF) for textual feature representation and Cosine Similarity for measuring similarities between tourism destinations. The research consists of several stages, including dataset collection, data loading, preprocessing, vectorization, similarity calculation, recommendation generation, and evaluation, as illustrated in Figure 1.

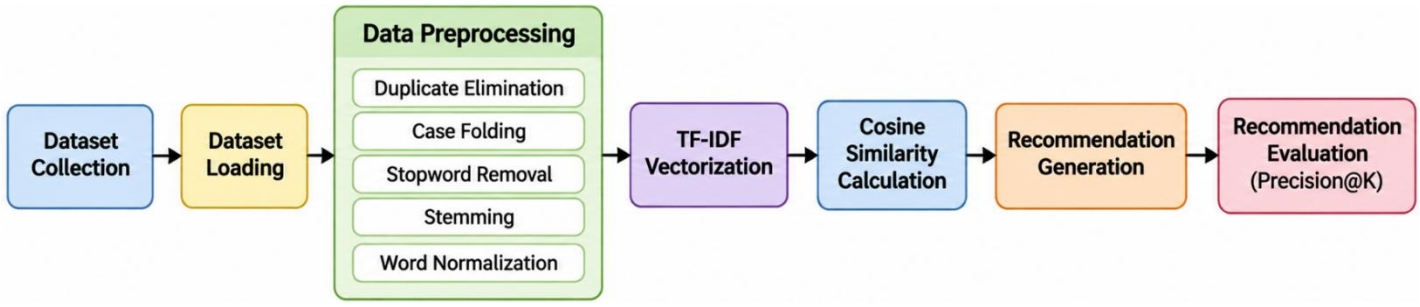

Figure 1. Research Workflow of the Tourism Recommendation System

a. Dataset Collection and Loading

The dataset consists of two components: tourism destination information and user ratings. The tourism dataset contains 50 natural tourism destinations represented by four attributes: destination ID, destination name, category, and description. The descriptions serve as the primary input for textual feature extraction, while user ratings are used to evaluate recommendation relevance.

All datasets are stored in CSV format and imported using Python and the pandas library.

Initial data inspection is performed to verify attribute completeness, identify missing values, and ensure data consistency before further processing.

b. Text Preprocessing

Text preprocessing is performed to improve the quality and consistency of destination descriptions before vectorization. The procedures include duplicate elimination, case folding, stopword removal, stemming, and word normalization. Duplicate elimination removes redundant destination records, while case folding converts all text into lowercase. Stopword removal eliminates frequently occurring words with limited contextual significance, whereas stemming converts words into their root forms (Rizki et al., 2019; Rumaisa, 2025). Word normalization standardizes textual representations to ensure consistent feature extraction. The resulting descriptions are used as input for TF-IDF vectorization.

c. Document Vectorization Using TF-IDF

The preprocessed destination descriptions are transformed into numerical feature vectors using TF-IDF. This method assigns weights to terms based on their frequency within a document and their distribution across the entire dataset (Kim & Gil, 2019; Qaiser & Ali, 2018). The TF-IDF value is calculated as follows:

$$TFIDF(t,d) = TF(t,d)xIDF(t)$$

Where:

$$IDF(t) = log\left(\frac{N}{df_t}\right)$$

Where $f_{t,d}$ represents the frequency of term t in document $d, \sum k\, f_{k,d}$ represents the total term frequency in document d, N denotes the total number of documents, and df_t represents the number of documents containing term t.

The resulting TF-IDF matrix represents each tourism destination as a numerical vector, which is subsequently used for similarity calculation.

d. Similarity Calculation Using Cosine Similarity

Cosine Similarity is applied to measure the similarity between tourism destinations based on their TF-IDF vectors. This method calculates the cosine of the angle between two vectors, allowing textual descriptions with different lengths to be compared based on their feature distributions (Melyani, 2022; Wang & Dong, 2020). The similarity is calculated using the following equation:

$$CosineSimilarity(d_1, d_2) = \frac{\sum_{i=1}^{n}(w_i d_1\, x\, w_i d_2)}{\sqrt{\sum_{i=1}^{n} w_{i,d_1}^2}\, x\, \sqrt{\sum_{i=1}^{n} w_{i,d_2}^2}}$$

Where $w_i d_1$ and $w_i d_2$ represent the TF-IDF weights of term i in destinations d_1 and d_2, respectively, and n denotes the total number of terms in the vocabulary.

The resulting similarity scores range from 0 to 1, with higher values indicating greater similarity between destination descriptions. These scores are organized into a similarity matrix for generating recommendations.

e. Recommendation Generation

The recommendation process begins when a user selects a reference destination. The system retrieves its similarity scores against other destinations and ranks them in descending order (Ko et al., 2022; Minarni & Sigit, 2023). The selected destination is excluded from the results, and the Top-K destinations with the highest similarity scores are presented as recommendations. The system displays destination names, similarity scores, ratings, and ticket prices to support users in selecting relevant tourism alternatives.

f. Recommendation Evaluation Using Precision@K

The recommendation system is evaluated using Precision@K to measure the proportion of relevant destinations among the generated recommendations (Silveira et al., 2019; Zangerle & Bauer, 2022). A destination is considered relevant when its user rating exceeds 4 on a five-point scale. Precision@K is calculated as follows:

$$Precision@K = \frac{Number of Relevant Recommendations}{K}$$

Where K represents the number of recommended destinations.

The evaluation is conducted using several reference destinations, and the resulting precision values are averaged to obtain the overall performance:

$$AveragePrecision@K = \frac{1}{M}\sum_{j=1}^{M} Precision@K_j$$

Where M represents the total number of evaluation scenarios and $Precision@K_j$ denotes the precision value obtained from each scenario. The average value indicates the system's ability to generate relevant tourism recommendations based on destination characteristics and user ratings.

Results and Discussion

This section presents the results of the implementation and evaluation of the proposed Content-Based Filtering approach for natural tourism destination recommendations in Labuan Bajo. The discussion focuses on examining the effectiveness of textual feature representation using TF-IDF, analyzing destination similarities through Cosine Similarity, and evaluating the recommendation system's performance through Precision@K and its web-based implementation. These discussions are organized into three main aspects to provide a comprehensive understanding of the proposed system, from destination feature representation to recommendation generation and performance evaluation.

a. Textual Feature Representation of Tourism Destinations Using TF-IDF

The quality of textual feature representation plays an important role in determining the effectiveness of content-based recommendation systems. In this study, tourism destination descriptions are utilized as the primary source of information to identify distinctive characteristics and establish relationships between destinations. The dataset comprises 50 natural tourism destinations in Labuan Bajo, with each destination represented by four attributes: identification number, destination name, tourism category, and textual description. The descriptions contain information about geographical characteristics, natural landscapes, environmental conditions, and tourism activities. These textual attributes provide the basis for

identifying similarities between destinations through the TF-IDF weighting mechanism. The use of weighted terms is consistent with the principles of information retrieval, in which term importance contributes to the effectiveness of document representation and comparison (Kim & Gil, 2019; Qaiser & Ali, 2018).

Table 1. Sample of Natural Tourism Destination Dataset

Destination	Category	Description
Komodo Island	Island	Island with natural Komodo habitat and beautiful natural and marine panoramas
Rinca Island	Island	Conservation island with trekking trails and the presence of wild animals
Pink Beach	Beach	Beach with pink-colored sand and clear seawater
Bidadari Island	Island	Small island with white sand, sea views, and natural panoramas from a height

Table 1 illustrates the diversity of natural tourism characteristics represented in the dataset. Komodo Island and Rinca Island share the island category but emphasize different descriptive features, including natural wildlife habitats, conservation activities, and trekking experiences. In contrast, Pink Beach is characterized by its distinctive pink-colored sand and clear seawater, representing a different tourism experience. Bidadari Island emphasizes white sandy beaches and panoramic natural landscapes. These variations demonstrate that destination classification alone is insufficient to represent the specific characteristics of tourism attractions. Therefore, textual descriptions provide more detailed information for distinguishing destinations within and across tourism categories. The 50-destination dataset provides the textual foundation for subsequent preprocessing and feature extraction.

The preprocessing stage is subsequently applied to improve textual consistency and reduce irrelevant information before feature extraction. The process includes text normalization, punctuation removal, stopword removal, and stemming, as described in the experimental results. These operations are intended to retain meaningful terms while reducing textual variations that may interfere with the similarity calculation. In particular, stopword removal eliminates frequently occurring words with limited descriptive significance, whereas stemming transforms derived words into their root forms. This approach supports more consistent term representation and reduces unnecessary textual features in the document collection (HaCohen-Kerner et al., 2020; Rizki et al., 2019).

For example, the description of Komodo Island, originally written as "Pulau dengan habitat asli komodo dan panorama alam", is transformed into "pulau habitat komodo panorama alam". Similarly, the description of Pink Beach is simplified by removing less relevant words and reducing derived terms into their root forms. These transformations preserve important descriptive characteristics while producing more concise textual representations. The results of the preprocessing stage are illustrated in Table 2.

Table 2. Comparison of Tourism Descriptions Before and After Preprocessing

Destination	Original Description	Main Preprocessing Transformation	Preprocessed Description
Komodo Island	Pulau dengan habitat asli komodo dan panorama alam	Stopword removal and text normalization	pulau habitat komodo panorama alam
Pink Beach	Pantai berpasir merah muda dengan laut jernih	Stemming and stopword removal	pantai pasir merah muda laut jernih

Destination	Original Description	Main Preprocessing Transformation	Preprocessed Description
Padar Island	Pulau berbukit dengan pemandangan laut	Stemming and stopword removal	pulau bukit pandang laut

The results presented in Table 2 demonstrate how preprocessing transforms the original descriptions into more concise textual representations while retaining their distinctive characteristics. For Komodo Island, irrelevant words such as dengan, asli, and dan are eliminated, leaving terms that represent its natural habitat and landscape. In the Pink Beach description, the word berpasir is reduced to pasir, while the distinctive expression merah muda is retained to represent the beach's unique physical appearance. Similarly, the description of Padar Island is simplified by reducing berbukit to bukit and pemandangan to pandang. These transformations improve textual consistency and help the TF-IDF process focus on descriptive terms that distinguish one destination from another.

Following preprocessing, the cleaned descriptions are transformed into numerical feature representations using the Term Frequency–Inverse Document Frequency (TF-IDF) method. TF-IDF assigns weights to individual terms by considering their frequency within a particular destination description and their distribution across the entire dataset. Terms that frequently occur in a specific description but appear less frequently in other descriptions receive relatively higher weights. This weighting mechanism helps distinguish the characteristic features of individual destinations and provides a quantitative basis for comparing their textual content. Chen (2024) likewise shows that refinements of TF-IDF term weighting can improve the identification of representative terms within a document collection. In this study, the resulting TF-IDF matrix represents each tourism destination as a numerical vector containing weighted descriptive terms.

Table 3. TF-IDF Feature Representation of Selected Tourism Destinations

Destination	Island	Beach	Sea	Hill	Komodo
Komodo Island	0.46	0.00	0.31	0.00	0.61
Rinca Island	0.44	0.00	0.29	0.00	0.57
Pink Beach	0.00	0.54	0.48	0.00	0.00
Padar Island	0.40	0.00	0.35	0.52	0.00

Table 3 presents the TF-IDF feature weights for four selected tourism destinations using five representative terms. The results indicate that each destination exhibits a different distribution of textual feature weights. Komodo Island records its highest weight for the term komodo (0.61), reflecting its distinctive association with the natural habitat of Komodo dragons. Rinca Island also demonstrates a relatively high weight for the same term (0.57), accompanied by weights for island (0.44) and sea (0.29). In contrast, Pink Beach has higher weights for beach (0.54) and sea (0.48), reflecting its coastal characteristics. Padar Island exhibits a different feature distribution, with a relatively high weight for hill (0.52), representing its distinctive hilly landscape. These results indicate that TF-IDF can differentiate destination characteristics through the relative importance of descriptive terms rather than relying solely on general tourism categories.

To provide a clearer understanding of the transformation from textual descriptions into numerical features, the relationship between

preprocessing and TF-IDF vectorization can be illustrated through a conceptual representation of the feature extraction process.

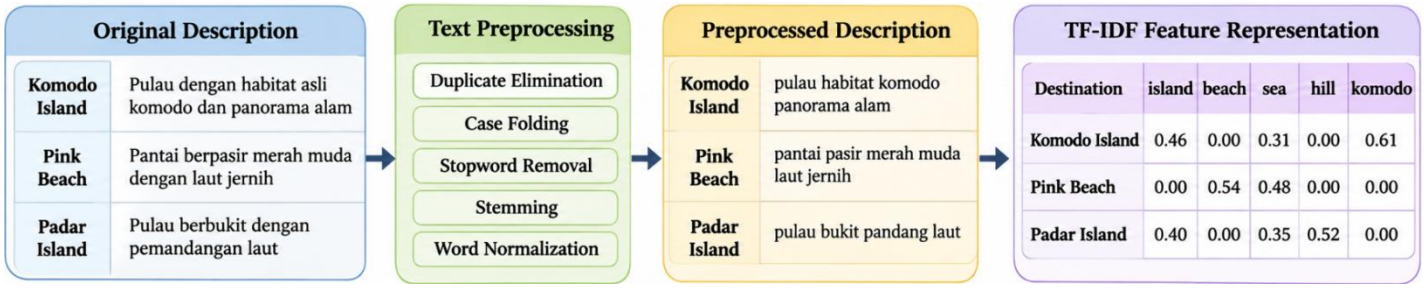

Figure 2. Transformation of Tourism Destination Descriptions into TF-IDF Feature Representations

The overall findings demonstrate that textual feature representation is an essential component of the proposed Content-Based Filtering system. The preprocessing results show that destination descriptions can be standardized while preserving their distinctive information, whereas TF-IDF converts these descriptions into numerical vectors that capture the relative importance of individual terms. The different weight distributions observed among Komodo Island, Rinca Island, Pink Beach, and Padar Island indicate that the system can represent variations in destination characteristics at the textual feature level. This representation subsequently provides the computational foundation for Cosine Similarity to identify relationships between destinations and generate content-based tourism recommendations. Nevertheless, TF-IDF primarily captures term frequency and distribution rather than contextual meaning or semantic relationships between words (Kusumaningrum & Rohman, 2026; Wang, 2024). Therefore, the quality of the resulting recommendations remains dependent on the completeness and descriptive accuracy of the available destination information.

b. Similarity Analysis Between Tourism Destinations Using Cosine Similarity

Following the TF-IDF vectorization process, Cosine Similarity is applied to identify the degree of similarity between natural tourism destinations based on their textual feature representations. Each destination is represented as a numerical vector containing weighted terms extracted from its description. The similarity calculation focuses on the angular relationship between two vectors, allowing destinations with different descriptive lengths to be compared based on their shared textual characteristics. The resulting similarity scores range from 0 to 1, where values closer to 1 indicate stronger textual similarity, while values closer to 0 indicate weaker similarity.

In this study, the similarity calculation is performed using the TF-IDF matrix generated from 50 natural tourism destinations in Labuan Bajo. The resulting similarity matrix provides a quantitative representation of relationships between destinations and serves as the basis for identifying relevant alternatives. This approach is consistent with the application of similarity-based recommendation techniques in tourism information systems, where destination characteristics and user requirements are considered in generating recommendations (Pasaribu & Sitompul, 2023; Pratiwi et al., 2025).

Table 4. Cosine Similarity Matrix Between Selected Tourism Destinations

Destination	Komodo Island	Rinca Island	Pink Beach	Padar Island
Komodo Island	1.00	0.91	0.19	0.42
Rinca Island	0.91	1.00	0.17	0.39
Pink Beach	0.19	0.17	1.00	0.23
Padar Island	0.42	0.39	0.23	1.00

Table 4 presents the Cosine Similarity matrix for four selected tourism destinations. The diagonal values of 1.00 indicate that each destination has complete similarity with itself, whereas the off-diagonal values represent the similarity between different destinations. The highest similarity score is observed between Komodo Island and Rinca Island (0.91), indicating substantial overlap in their textual characteristics. Meanwhile, Komodo Island and Padar Island demonstrate a moderate similarity score of 0.42, followed by Rinca Island and Padar Island at 0.39. In contrast, Pink Beach exhibits relatively low similarity with the other destinations, with scores ranging from 0.17 to 0.23. These findings demonstrate that the algorithm can distinguish destinations based on descriptive characteristics rather than relying exclusively on their general tourism classifications.

The similarity patterns observed in Table 4 can be further examined by comparing individual destination pairs and their corresponding descriptive characteristics. Such comparisons are important because numerical similarity scores provide quantitative evidence, while examining the underlying textual features helps explain why particular destinations are identified as similar or different.

Table 5. Pairwise Similarity Analysis Based on Destination Characteristics

Destination Pair	Cosine Similarity	Shared Characteristics	Distinguishing Characteristics
Komodo Island – Rinca Island	0.91	Island, wildlife, conservation	Komodo habitat and trekking activities
Komodo Island – Padar Island	0.42	Island, sea, natural landscape	Wildlife conservation versus panoramic scenery
Rinca Island – Padar Island	0.39	Island, sea, natural landscape	Wildlife observation versus hilly geographical features
Pink Beach – Padar Island	0.23	Sea and natural landscape	Pink sandy beach versus hilly island landscape
Komodo Island – Pink Beach	0.19	Sea and natural environment	Wildlife conservation versus coastal recreation
Rinca Island – Pink Beach	0.17	Sea and natural environment	Wildlife conservation versus beach activities

Table 5 provides a comparative interpretation of the similarity relationships among the selected destinations. The strongest relationship occurs between Komodo Island and Rinca Island, with a score of 0.91. This result corresponds to their shared characteristics as islands associated with wildlife conservation and natural habitats. In contrast, Padar Island demonstrates lower similarity with both destinations, despite belonging to the same general island category. Its descriptive emphasis on hills, geographical formations, and panoramic scenery differs from the conservation-oriented characteristics of Komodo and Rinca Islands. Pink Beach exhibits

the lowest similarity values, particularly with Rinca Island (0.17) and Komodo Island (0.19). This distinction reflects the different descriptive focus of coastal recreation compared with wildlife conservation. The findings indicate that textual feature comparison provides a more detailed distinction between destinations than general category classification alone. This interpretation is consistent with previous research highlighting the importance of destination characteristics in tourism recommendation processes (Hamid et al., 2021; Masmur et al., 2026).

Furthermore, the similarity scores can be arranged in descending order to identify the strongest relationships between different tourism destinations. This ranking provides an additional perspective on how the proposed method prioritizes destination alternatives based on their textual characteristics.

Table 6. Ranking of Similarity Scores Between Tourism Destination Pairs

Rank	Destination Pair	Cosine Similarity	Similarity Interpretation
1	Komodo Island – Rinca Island	0.91	Highest similarity
2	Komodo Island – Padar Island	0.42	Moderate similarity
3	Rinca Island – Padar Island	0.39	Moderate similarity
4	Pink Beach – Padar Island	0.23	Low similarity
5	Komodo Island – Pink Beach	0.19	Low similarity
6	Rinca Island – Pink Beach	0.17	Lowest similarity

Table 6 presents the ranking of six unique destination pairs based on their Cosine Similarity scores. Komodo Island and Rinca Island occupy the highest position, with a similarity score of 0.91, indicating a strong correspondence between their textual descriptions. The second and third positions are occupied by Komodo Island–Padar Island (0.42) and Rinca Island–Padar Island (0.39), respectively. These moderate scores suggest that although the destinations share certain geographical characteristics, their descriptive features emphasize different tourism experiences. The remaining pairs exhibit relatively low similarity, particularly Rinca Island and Pink Beach, with a score of 0.17. This ranking demonstrates the system's ability to distinguish varying degrees of textual similarity and establish a quantitative basis for recommendation prioritization. It should be noted that the qualitative labels in this table describe the relative ordering of the observed scores rather than predefined universal similarity thresholds.

To further illustrate the distribution of similarity scores, a visual representation of the relationships between selected tourism destinations can be presented. This visualization is intended to emphasize the strongest and weakest similarity relationships identified through the Cosine Similarity calculation.



Figure 3. Heatmap Visualization of Cosine Similarity Between Selected Natural Tourism Destinations in Labuan Bajo

The overall findings demonstrate that Cosine Similarity can identify meaningful relationships between tourism destinations based on their textual feature representations. The high similarity between Komodo Island and Rinca Island reflects their overlapping conservation-related characteristics, whereas the lower similarity values involving Pink Beach indicate differences in their descriptive features and tourism experiences. Furthermore, the moderate similarity values involving Padar Island demonstrate that destinations within the same broad category may still exhibit distinct textual characteristics. These results suggest that combining TF-IDF and Cosine Similarity provides a systematic mechanism for identifying destination relationships beyond conventional category-based classification. The resulting similarity matrix can therefore support the recommendation process by prioritizing destinations with stronger descriptive correspondence to a user's selected reference destination. Nevertheless, Cosine Similarity measures textual feature overlap rather than directly measuring actual tourist satisfaction or behavioral preference (Zangerle & Bauer, 2022). Consequently, the similarity scores should be interpreted as indicators of content similarity rather than definitive evidence of user preference.

c. Recommendation Performance Evaluation and System Implementation

The performance of the proposed tourism recommendation system is evaluated using Precision@K to assess the relevance of the recommended destinations. In this study, Precision@5 is employed to measure the proportion of relevant destinations among the five highest-ranked recommendations generated for each selected reference destination. A recommendation is classified as relevant when the corresponding user rating exceeds 4 on a five-point scale. This evaluation approach provides an indication of how effectively the system identifies destinations that satisfy the

predefined relevance criteria. Precision is commonly used in recommender system evaluation to assess the proportion of retrieved items that are relevant to users (Herlawati et al., 2026; Silveira et al., 2019).

The evaluation is conducted through five testing scenarios using different reference destinations: Komodo Island, Pink Beach, Padar Island, Rinca Island, and Bidadari Island. In each scenario, the system generates five recommendations based on the highest Cosine Similarity scores. The generated recommendations are subsequently compared with the user rating dataset to determine the number of relevant results.

Table 7. Precision@5 Evaluation Results Across Different Reference Destinations

Test	Reference Destination	Total Recommendations	Relevant Recommendations	Precision@5
1	Komodo Island	5	4	0.80
2	Pink Beach	5	3	0.60
3	Padar Island	5	4	0.80
4	Rinca Island	5	5	1.00
5	Bidadari Island	5	4	0.80
Average	**All scenarios**	**5**	**4**	**0.80**

Table 7 presents the evaluation results across five reference destinations. The findings reveal that Rinca Island achieves the highest Precision@5 value of 1.00, indicating that all five recommended destinations satisfy the predefined relevance criterion. Komodo Island, Padar Island, and Bidadari Island each achieve a precision value of 0.80, meaning that four out of five recommendations are relevant. Meanwhile, Pink Beach records the lowest precision value of 0.60, with only three relevant recommendations. The overall average Precision@5 of 0.80 indicates that four out of every five recommendations are relevant across the five evaluation scenarios. These variations suggest that recommendation performance depends partly on the characteristics of the selected reference destination and the degree to which its textual features correspond to those of other destinations.

To further examine the evaluation results, the distribution of relevant and non-relevant recommendations can be analyzed across the five testing scenarios. This comparison provides a clearer understanding of the system's recommendation consistency and identifies destinations that require further improvement.

Table 8. Distribution of Relevant and Non-Relevant Recommendations

Reference Destination	Relevant	Non-Relevant	Relevance Rate (%)	Non-Relevance Rate (%)
Komodo Island	4	1	80	20
Pink Beach	3	2	60	40
Padar Island	4	1	80	20
Rinca Island	5	0	100	0
Bidadari Island	4	1	80	20
Total	**20**	**5**	**80**	**20**

Table 8 provides a more detailed comparison of recommendation relevance across the evaluation scenarios. Overall, the system generates 20 relevant recommendations and five non-relevant recommendations from 25 recommendation results, producing an aggregate relevance rate of 80%. Rinca Island demonstrates the strongest performance, with no non-relevant recommendations. In contrast, Pink Beach produces two non-relevant recommendations, representing 40% of its results. This variation indicates that the system performs more consistently when destination descriptions contain distinctive characteristics that are also represented in other relevant destinations. However, destinations with diverse recreational activities may require more detailed textual attributes to distinguish user preferences more accurately (Aryanto et al., 2025; Maulana et al., 2026). These findings highlight the importance of descriptive feature quality in determining recommendation relevance.

The differences in evaluation results can also be interpreted by examining the relationship between destination characteristics and recommendation performance. Rinca Island achieves perfect precision, while Pink Beach exhibits comparatively lower performance. According to the original experimental discussion, the latter limitation is associated with differences in beach-related activities, such as snorkeling and sunbathing, which may not be adequately distinguished through conventional term-based representations. This observation reflects a limitation of TF-IDF, which primarily captures the statistical importance of terms rather than their contextual or semantic relationships. Previous studies have also identified the importance of incorporating richer item characteristics and contextual information to improve recommendation relevance (Javed et al., 2021; Renjith et al., 2020).

Table 9. System Implementation and Recommendation Features

System Component	Implementation	Main Functionality	User Benefit
Backend	Python	Executes Content-Based Filtering and Cosine Similarity calculations	Supports recommendation processing
Frontend	HTML, CSS, and JavaScript	Provides an interactive web interface	Facilitates user interaction
Destination Selection	Destination selection menu	Allows users to select a reference destination	Supports personalized exploration
Recommendation Output	Top-5 recommendation list	Displays destinations ranked by similarity	Helps users compare alternatives
Similarity Information	Cosine Similarity score (0–1)	Indicates the degree of textual similarity	Improves recommendation transparency
Destination Information	Rating and ticket price	Presents additional destination information	Supports destination selection
Dynamic Exploration	Interactive recommendation selection	Allows a recommended destination to become a new reference	Enables continuous destination exploration

Table 9 summarizes the main components and functionalities of the implemented tourism recommendation application. The backend, developed using Python, performs the recommendation calculations, while HTML, CSS, and JavaScript provide the interactive frontend. Users can select a reference destination and obtain five recommended alternatives ranked according to their similarity scores. The interface also presents destination ratings and ticket prices to provide additional information for decision-making. One notable feature is dynamic exploration, which allows users to select a recommended destination as a new reference point and obtain another recommendation list. This functionality supports continuous exploration of alternative destinations without requiring users to restart the application. The system reportedly generates recommendations in less than one second and supports access through desktop, tablet, and smartphone devices, as described in the original implementation results.

The system implementation further demonstrates how the proposed recommendation approach can be translated into a practical application for tourism exploration (Fakhri & Nugroho, 2025; Kontogianni & Alepis, 2020). The web-based interface integrates destination selection, similarity-based ranking, and destination information into a single environment (Nurpila et al., 2024; Prasetyo et al., 2021). Rather than presenting similarity scores as isolated numerical outputs, the application provides users with contextual information that can support destination comparisons (Zhang & Chen, 2020).

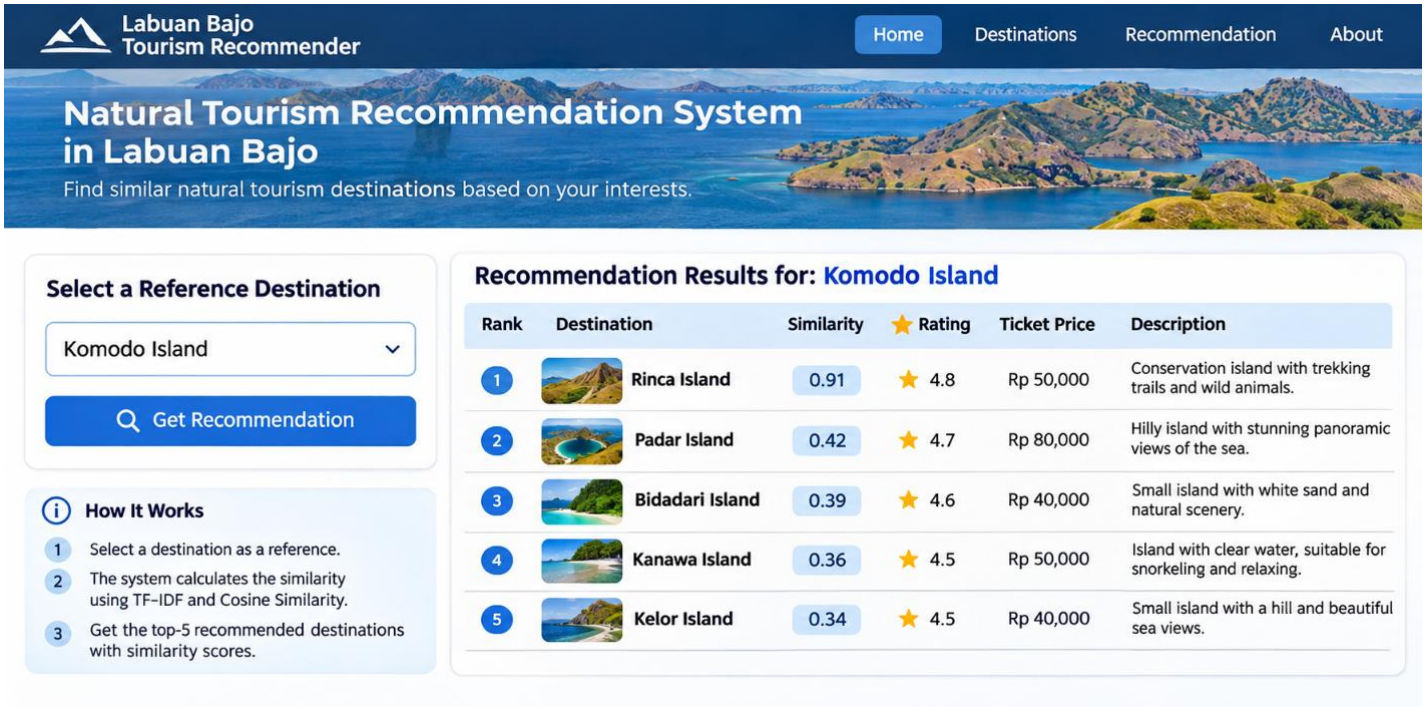

Figure 4. Web-Based Interface of the Labuan Bajo Natural Tourism Recommendation System

Figure 4 is intended to illustrate the actual system interface, including the destination selection menu, recommendation configuration, and ranked destination results. The interface also demonstrates how similarity scores and destination information are presented to users. Its dynamic interaction feature allows users to select recommended destinations as new reference points, thereby supporting continuous exploration of tourism alternatives.

Overall, the evaluation results indicate that the proposed Content-Based Filtering system achieves an average Precision@5 of 0.80 across five testing scenarios. This result demonstrates that the integration of TF-IDF and Cosine Similarity can generate relevant tourism recommendations based on destination descriptions. However, the variation between Rinca Island and Pink Beach highlights the limitations of relying exclusively on statistical term representations. In particular, destinations with similar general characteristics may offer different recreational experiences that are not fully captured through textual feature overlap. Therefore, incorporating more detailed destination attributes, activity classifications, or semantic text representations could provide opportunities for future improvements (Kusumaningrum & Rohman, 2026; Zhang et al., 2019). From a practical perspective, the developed web application offers an accessible mechanism for exploring natural tourism alternatives in Labuan Bajo. Nevertheless, the observed precision should be interpreted within the scope of the five evaluation scenarios and the rating-based relevance criterion rather than as a comprehensive measure of overall user satisfaction; rank-sensitive and beyond-accuracy measures could complement this assessment (Silveira et al., 2019; Zangerle & Bauer, 2022). The findings consequently establish an empirical foundation for further development of personalized tourism recommendation systems.

Conclusion

This study successfully developed a natural tourism recommendation system for Labuan Bajo using an item-based Content-Based Filtering approach integrating TF-IDF and Cosine Similarity. The proposed approach represents tourism destination descriptions as numerical feature vectors, enabling the identification of similarities based on their distinctive textual characteristics. The experimental results demonstrate that TF-IDF can distinguish destination features, while Cosine Similarity effectively identifies relationships between destinations. The highest similarity score of 0.91 was obtained between Komodo Island and Rinca Island, indicating substantial similarities in their descriptive characteristics. In contrast, Pink Beach exhibited relatively low similarity scores with other destinations, reflecting its distinctive coastal characteristics.

The recommendation system was evaluated using Precision@5 across five testing scenarios, resulting in an average precision of 0.80. Rinca Island achieved the highest precision value of 1.00, whereas Pink Beach recorded the lowest value of 0.60. These findings indicate that the proposed approach can generate relevant tourism recommendations, although its performance varies depending on the descriptive characteristics of each destination. The system was also implemented as an interactive web application that enables users to select reference destinations, explore ranked recommendations, and access relevant destination information.

Overall, this study demonstrates the practical applicability of integrating TF-IDF and Cosine Similarity to support personalized natural tourism exploration in Labuan Bajo. Future research should incorporate richer destination attributes, contextual semantic representations, and more comprehensive user preference data to improve recommendation relevance and evaluation reliability.

References

Albab, J. F. U., & Rohman, A. N. (2026). Recommendation system Yogyakarta tourism using TF-IDF and cosine similarity methods with word normalizer. *Journal of Applied Informatics and Computing, 10*(1), 935–945. https://doi.org/10.30871/jaic.v10i1.11751

Al Fararni, K., Nafis, F., Aghoutane, B., Yahyaouy, A., Riffi, J., & Sabri, A. (2021). Hybrid recommender system for tourism based on big data and AI: A conceptual framework. *Big Data Mining and Analytics, 4*(1), 47–55. https://doi.org/10.26599/BDMA.2020.9020015

Alyari, F., & Jafari Navimipour, N. (2018). Recommender systems: A systematic review of the state of the art literature and suggestions for future research. *Kybernetes, 47*(5), 985–1017. https://doi.org/10.1108/K-06-2017-0196

Ardin, A. (2024a, January 7). Lampaui target, 423.847 turis kunjungi Labuan Bajo sepanjang 2023. *detikBali*. https://www.detik.com/bali/wisata/d-7129064/lampaui-

Ardin, A. (2024b, January 4). Naik drastis! 300.488 turis kunjungi TN Komodo sepanjang 2023. *detikBali*. https://www.detik.com/bali/wisata/d-7123901/naik-drastis-300-488-turis-kunjungi-tn-komodo-sepanjang-2023

Aryanto, A. D., Primadewi, A., & Prabowo, N. A. (2025). Rekomendasi wisata Kabupaten Magelang menggunakan metode content-based filtering dan location-based service. *Jurnal Fasilkom, 15*(1), 172–178. https://doi.org/10.37859/jf.v15i1.8156

Chen, L.-C. (2024). An extended TF-IDF method for improving keyword extraction in traditional corpus-based research: An example of a climate change corpus. *Data & Knowledge Engineering, 153*, Article 102322. https://doi.org/10.1016/j.datak.2024.102322

Fakhri, & Nugroho, K. (2025). Sistem rekomendasi wisata di Pekalongan melalui chatbot dengan framework Rasa. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), 9*(1), 68–77. https://doi.org/10.35870/jtik.v9i1.3000

Faurina, R., & Sitanggang, E. (2023). Implementasi metode content-based filtering dan collaborative filtering pada sistem rekomendasi wisata di Bali. *Techno.Com, 22*(4), 870–881. https://doi.org/10.33633/tc.v22i4.8556

HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLOS ONE, 15*(5), Article e0232525. https://doi.org/10.1371/journal.pone.0232525

Hamid, R. A., Albahri, A. S., Alwan, J. K., Al-qaysi, Z. T., Albahri, O. S., Zaidan, A. A., Alnoor, A., Alamoodi, A. H., & Zaidan, B. B. (2021). How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management. *Computer Science Review, 39*, Article 100337. https://doi.org/10.1016/j.cosrev.2020.100337

Herlawati, H., Hendharsetiawan, A. A., Uskono, B. M., & Acah, A. (2026). Evaluasi kinerja

metode content-based filtering pada sistem rekomendasi destinasi wisata dan rute pendakian berbasis web. *Journal of Students' Research in Computer Science, 7*(1), 85–98. https://doi.org/10.31599/thhffe58

Huda, C., Heryadi, Y., Lukas, & Budiharto, W. (2024). Smart tourism recommender system modeling based on hybrid technique and content boosted collaborative filtering. *IEEE Access, 12*, 131794–131808. https://doi.org/10.1109/ACCESS.2024.3450882

Huda, C., Ramadhan, A., Trisetyarso, A., Abdurachman, E., & Heryadi, Y. (2021). Smart tourism recommendation model: A systematic literature review. *International Journal of Advanced Computer Science and Applications, 12*(12). https://doi.org/10.14569/IJACSA.2021.0121222

Islahuddin, Eppang, B. M., Som, A. P. M., Masatip, A., & Salim, M. A. M. (2022). Adaptation and collaboration of local community in super priority destination tourism program in Labuan Bajo. *Etnosia: Jurnal Etnografi Indonesia, 7*(1), 15–28. https://doi.org/10.31947/etnosia.v7i1.19708

Javed, U., Shaukat, K., Hameed, I. A., Iqbal, F., Alam, T. M., & Luo, S. (2021). A review of content-based and context-based recommendation systems. *International Journal of Emerging Technologies in Learning, 16*(3), 274–306. https://doi.org/10.3991/ijet.v16i03.18851

Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review, 52*(1), 273–292. https://doi.org/10.1007/s10462-018-09677-1

Kim, S.-W., & Gil, J.-M. (2019). Research paper classification systems based on TF-IDF and LDA schemes. *Human-centric Computing and Information Sciences, 9*, Article 30. https://doi.org/10.1186/s13673-019-0192-7

Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A survey of recommendation systems: Recommendation models, techniques, and application fields. *Electronics, 11*(1), Article 141. https://doi.org/10.3390/electronics11010141

Kontogianni, A., & Alepis, E. (2020). Smart tourism: State of the art and literature review for the last six years. *Array, 6*, Article 100020. https://doi.org/10.1016/j.array.2020.100020

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information, 10*(4), Article 150. https://doi.org/10.3390/info10040150

Kusumaningrum, L. B., & Rohman, A. N. (2026). Tourism recommendation system for Labuan Bajo using hybrid TF-IDF and Sentence-BERT with semantic feature boosting. *Journal of Applied Informatics and Computing, 10*(5), 4670–4681. https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/13109

Lops, P., Jannach, D., Musto, C., Bogers, T., & Koolen, M. (2019). Trends in content-based recommendation. *User Modeling and User-Adapted Interaction, 29*(2), 239–249. https://doi.org/10.1007/s11257-019-09231-w

Masmur, S. S., Fajri, I. N., & Rohman, A. N. (2026). Sistem rekomendasi wisata kuliner Labuan Bajo dengan perbandingan metode similarity content-based filtering. *Sistemasi: Jurnal Sistem Informasi, 15*(8), 3287–3302. https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/6830

Maulana, A., Rohman, A. N., & Pristyanto, Y. (2026). Sistem rekomendasi destinasi tempat wisata Provinsi Aceh menggunakan metode content based filtering. *Sistemasi: Jurnal*

Sistem Informasi, 15(2), 418–432. https://doi.org/10.32520/stmsi.v15i2.5500

Melyani, C. A. (2022). Hotel recommendation system with content-based filtering approach (Case study: Hotel in Yogyakarta on Nusatrip website). *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika, 15*(1). https://doi.org/10.36456/jstat.vol15.no1.a5375

Minarni, M., & Sigit, S. (2023). Pengembangan sistem informasi rekomendasi wisata Kotawaringin Timur berbasis web menggunakan metode item-based collaborative filtering. *Jurnal Ilmiah Informatika Global, 13*(3). https://doi.org/10.36982/jiig.v13i3.2695

Nurpila, S., Ismail, H. A., & Prakoso, S. A. (2024). Rancang bangun rekomendasi tempat wisata di Kabupaten Rembang berbasis website menggunakan metode content based filtering. *INTECOMS: Journal of Information Technology and Computer Science, 7*(4), 1149–1159.

Oyadila, S., Abdullah, D., & Razi, A. (2025). Implementasi content-based filtering dengan TF-IDF dan cosine similarity untuk sistem rekomendasi destinasi wisata di Aceh Tengah. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab, 10*(2), 1329–1339. https://doi.org/10.36341/rabit.v10i2.6532

Pahlevi, M. R. (2024). Human rights violations in the development of the super priority tourism destination Labuan Bajo. *Jurnal Pemikiran Sosiologi, 11*(1), 26–48. https://doi.org/10.22146/jps.v11i1.97299

Panda, D. K., & Ray, S. (2022). Approaches and algorithms to mitigate cold start problems in recommender systems: A systematic literature review. *Journal of Intelligent Information Systems, 59*(2), 341–366. https://doi.org/10.1007/s10844-022-00698-5

Panjaitan, C. H. P., Pangaribuan, L. J., & Cahyadi, C. I. (2022). Analisis metode k-nearest neighbor menggunakan Rapid Miner untuk sistem rekomendasi tempat wisata Labuan Bajo. *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer, 6*(3), 534–541. https://doi.org/10.33395/remik.v6i3.11701

Pasaribu, Y. S., & Sitompul, T. S. (2023). Rekomendasi destinasi wisata Kota Bandung menggunakan algoritma collaborative filtering. *Mutiara: Jurnal Penelitian dan Karya Ilmiah, 1*(6), 382–392. https://doi.org/10.59059/mutiara.v1i6.736

Prasetyo, B., Atina, V., & Purwanto, E. (2021). Sistem rekomendasi pariwisata dengan metode content based recommendation berbasis website (Studi kasus: Dinas Pariwisata dan Budaya Surakarta). *DutaCom, 14*(1), 51–58. https://doi.org/10.47701/dutacom.v14i1.2017

Pratiwi, D., Asrianda, A., & Rosnita, L. (2025). Penerapan metode content-based filtering dalam sistem rekomendasi objek wisata di Aceh Tamiang. *Jurnal Ilmu Komputer dan Informatika, 4*(2), 85–96. https://doi.org/10.54082/jiki.169

Qaiser, S., & Ali, R. (2018). Text mining: Use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications, 181*(1), 25–29. https://doi.org/10.5120/ijca2018917395

Renjith, S., Sreekumar, A., & Jathavedan, M. (2020). An extensive study on the evolution of context-aware personalized travel recommender systems. *Information Processing & Management, 57*(1), Article 102078. https://doi.org/10.1016/j.ipm.2019.102078

Rizki, A. S., Tjahyanto, A., & Trialih, R. (2019). Comparison of stemming algorithms on Indonesian text processing. *TELKOMNIKA (Telecommunication Computing Electronics*

and Control), 17(1), 95–102. https://doi.org/10.12928/telkomnika.v17i1.10183

Roetzel, P. G. (2019). Information overload in the information age: A review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. *Business Research, 12*(2), 479–522. https://doi.org/10.1007/s40685-018-0069-z

Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data, 9*, Article 59. https://doi.org/10.1186/s40537-022-00592-5

Rumaisa, F. (2025). Evaluation of Indonesian language stemmer algorithms: A comparative analysis. *Brilliance: Research of Artificial Intelligence, 5*(1), 21–25. https://doi.org/10.47709/brilliance.v5i1.5679

Salsabilla, C., & Utomo, D. W. (2025). Sistem rekomendasi wisata Kabupaten Pekalongan dengan content based filtering. *Sistemasi: Jurnal Sistem Informasi, 14*(1), 262–270. https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/4839

Silveira, T., Zhang, M., Lin, X., Liu, Y., & Ma, S. (2019). How good your recommender system is? A survey on evaluations in recommendation. *International Journal of Machine Learning and Cybernetics, 10*(5), 813–831. https://doi.org/10.1007/s13042-017-0762-9

Suharyanto, Fauzi, A., Adrianto, L., & Wahyudin, Y. (2024). Developing marine tourism policy pathways: The case of super priority marine tourism destinations in Labuan Bajo, Indonesia. *International Journal of Sustainable Development and Planning, 19*(10), 3809–3819. https://doi.org/10.18280/ijsdp.191010

Tarigan, T. E., Faizal, E., & Sumiyatun. (2023). Model rekomendasi wisata dengan pendekatan collaborative filtering. *FAHMA: Jurnal Informatika Komputer, Bisnis dan Manajemen, 21*(2), 56–64. https://doi.org/10.61805/fahma.v21i2.18

Teguh, F. X., & Scott, N. (2026). Government-led DMO and collaborative governance practices: Case study in Labuan Bajo, Flores, Indonesia. *Jurnal Kepariwisataan Indonesia: Jurnal Penelitian dan Pengembangan Kepariwisataan Indonesia, 20*(1), 41–74. https://doi.org/10.47608/jki.v20i12026.41-74

Wang, J., & Dong, Y. (2020). Measurement of text similarity: A survey. *Information, 11*(9), Article 421. https://doi.org/10.3390/info11090421

Wang, Y. (2024). Research on the TF–IDF algorithm combined with semantics for automatic extraction of keywords from network news texts. *Journal of Intelligent Systems, 33*, Article 20230300. https://doi.org/10.1515/jisys-2023-0300

Yuan, H., & Hernandez, A. A. (2023). User cold start problem in recommendation systems: A systematic review. *IEEE Access, 11*, 136958–136977. https://doi.org/10.1109/ACCESS.2023.3338705

Zangerle, E., & Bauer, C. (2022). Evaluating recommender systems: Survey and framework. *ACM Computing Surveys, 55*(8), Article 170. https://doi.org/10.1145/3556536

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys, 52*(1), Article 5. https://doi.org/10.1145/3285029

Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. *Foundations and Trends in*

Information Retrieval, 14(1), 1–101. https://doi.org/10.1561/1500000066