BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS
Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.28

Content-Based Filtering with TF-IDF and Cosine Similarity for Bandung Raya Tourism Recommendation

Mira Aldina[1*], Galih[2], Siti Nur[3]

Informatics Engineering Study Program[1,2,3]
Universitas Islam Nusantara, Bandung, Indonesia[1,2,3]
www.uninus.ac.id[1,2,3]
miraaldina34@gmail.com[1*], galihsetiana@gmail.com[2], sitinuryendi@gmail.com[3]

Abstract. Bandung Raya offers hundreds of tourist destinations spread across Bandung City, Bandung Regency, and West Bandung Regency, which causes information overload and makes it difficult for tourists to select destinations that match their preferences. Recommender systems, as one of the most widely applied branches of machine learning, provide a way to filter such information automatically. This study develops a machine learning-based recommender system for tourism destinations in Bandung Raya using Content-Based Filtering, in which destination descriptions are represented as numerical vectors through TF-IDF weighting and compared using Cosine Similarity. The dataset consists of 331 destinations with name, description, category, and region attributes. Text preprocessing is performed in five stages using the Sastrawi library for the Indonesian language, producing a TF-IDF matrix of 331 by 583 and a similarity matrix of 331 by 331. The model is deployed as a website using Flask as the backend, React as the frontend, and a REST API as the interface, supporting both name-based search and free-text query search. Functional validation uses Black-Box Testing, while recommendation quality is measured using Precision at K and Mean Average Precision. All eight functional scenarios passed, with Precision at 5 of 96.00 percent, Precision at 10 of 90.00 percent, and Mean Average Precision of 98.86 percent, indicating that, for the five evaluated queries, relevant destinations are ranked highly.

Key words: machine learning; recommender system; content-based filtering; TF-IDF; cosine similarity

1. INTRODUCTION

Bandung Raya, comprising Bandung City, Bandung Regency, and West Bandung Regency, is one of the most visited tourism regions on Java Island. According to the Bandung City Culture and Tourism Office, the number of tourists in 2023 reached 7.7 million, an increase of approximately 17% from 6.6 million in the previous year, and continued to grow to 8.5 million in 2024 [1]. The region offers a wide variety of destinations, ranging from natural attractions to educational and culinary sites. This abundance of options, however, creates information overload, a condition in which users struggle to make decisions because the available information exceeds their processing capacity [2]. Tourists frequently spend a considerable amount of time searching through search engines, social media, and travel blogs, yet the results do not necessarily reflect their preferences.

It should be noted that, under Presidential Regulation (Peraturan Presiden) No. 45 of 2018 on the Spatial Plan of the Bandung Basin Urban Area [19], the wider Bandung metropolitan area (Cekungan Bandung) also designates Cimahi City as part of its urban core and includes several subdistricts of Sumedang Regency. This study, however, scopes its corpus to Bandung City, Bandung Regency, and West Bandung Regency, the administrative units for which consistent destination data could be assembled; the reported coverage claims are limited accordingly.

Digital platforms and mobile applications have been developed to improve the delivery of destination information to travellers [3]. Most of them, however, present static catalogues or rely on exact keyword matching, so the burden of filtering remains with the user: they do not model user preference and therefore cannot rank destinations by relevance. A recommender system addresses precisely this gap, acting as the intelligent component that decides which items a user is actually shown.

Recommender systems are among the most widely applied branches of machine learning and have been deployed extensively in e-commerce, streaming services, and tourism [4]. Two dominant paradigms exist: Collaborative Filtering (CF), which exploits interaction patterns among users, and Content-Based Filtering (CBF), which matches item attributes with user preferences [5]. CF requires a

Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.28

substantial volume of user interaction data and suffers from the cold-start problem, whereas CBF remains effective when interaction data are scarce, making it well suited to the tourism domain in which most destinations lack user ratings [6]. Web-based tourism recommenders built on this paradigm have been shown to be practical to deploy [7]. The collaborative and content-based paradigms were established in early work on information filtering and recommendation [20], [21] and have since been examined in numerous surveys of the field [22], [23], [24], [25].

This study therefore develops a machine learning-based recommender system for tourism destinations in Bandung Raya using Content-Based Filtering with TF-IDF and Cosine Similarity. The contributions are fourfold: (1) the construction of a curated corpus of 331 destinations in Bandung Raya; (2) an Indonesian-language preprocessing and TF-IDF pipeline based on the Sastrawi library; (3) the deployment of the trained model as a website built with Flask, React, and a REST API, providing both name-based and free-text query modes; and (4) an evaluation combining Black-Box Testing with Precision@K and Mean Average Precision (MAP).

2. RELATED WORK

Content-Based Filtering recommends items by analysing their attributes and matching them against a user profile or query. When those attributes are textual, the representation stage is an unsupervised learning problem: documents are mapped into a vector space in which semantically related items lie close to one another [5]. TF-IDF is the most widely used weighting scheme for this purpose, assigning a high weight to terms that occur frequently within a document but rarely across the corpus, while Cosine Similarity measures the angle between two vectors and is therefore insensitive to document length [8].

Content-based filtering has a long research history, ranging from early systems that recommend items by matching item attributes against a user profile [26], [27] to more recent network-based formulations that mitigate sparsity and over-specialisation [28]; general treatments of recommender-system techniques, including hybrid designs, are provided in [29], [30], [31].

Several studies have adopted this approach across different domains. Albab and Rohman [9] built a Yogyakarta tourism recommender using TF-IDF and Cosine Similarity, while Faurina and Sitanggang [6] combined Content-Based Filtering and Collaborative Filtering for Bali tourism. Salsabilla and Utomo [10] applied CBF to Pekalongan Regency, Wicaksono et al. [11] to café selection, and Timur [12] to Indonesian song recommendation, while Setiawan and Adnyana [13] employed the same representation in a helpdesk chatbot, demonstrating the versatility of the technique beyond recommendation. Alternative paradigms have also been explored: Pasaribu and Sitompul [14] used Collaborative Filtering for Bandung City tourism, and Saputra et al. [2] applied matrix factorisation to Indonesian tourist destinations.

These studies demonstrate the effectiveness of TF-IDF and Cosine Similarity for text-based recommendation, but they differ from the present work in three respects. First, none targets the Bandung Raya region with a corpus of comparable size. Second, most evaluate the model offline without deploying it behind an interactive interface. Third, few provide a free-text query mode that allows users to describe their preferences in natural language rather than selecting predefined categories. This study addresses these gaps.

3. METHODS

This study is applied research with a quantitative approach. System development uses the Waterfall model consisting of requirements analysis, design, implementation, testing, and maintenance [15]. The research stages are broadly shown in Figure 1, following an input–process–output flow: destination data is processed through text preprocessing, TF-IDF weighting, and Cosine Similarity computation, and then produces Top-K recommendations displayed through the website.

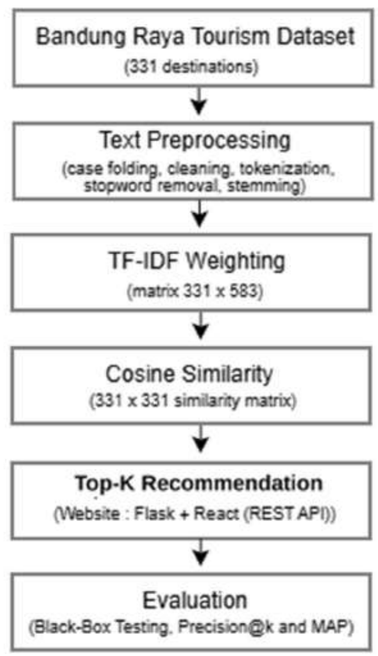

Fig.1. Research stages

3.1. Data Collection

The data used are secondary data on tourist destinations in the Bandung Raya region (Bandung City, Bandung Regency, and West Bandung Regency) obtained from tourism information websites, travel articles, and the official dataset of the West Java Provincial Tourism and Culture Office. After being selected based on regional relevance, description completeness, and removal of duplicate data,

BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS
Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.28

331 destinations were obtained. Each destination has four attributes, namely name, description, category, and region, with the description as the main attribute. The dataset is stored in CSV format.

In this study, description completeness was defined as the presence of a non-empty textual description of at least one sentence, and duplicate records were removed by retaining unique destination names and identifiers; the retained set was verified programmatically to contain exactly 331 records with 331 unique names and identifiers.

The dataset was compiled by the author as secondary data from three types of publicly available sources: (i) Indonesian tourism-information websites that publish destination profiles; (ii) travel articles and travel-blog write-ups; and (iii) the tourism-destination dataset made publicly available by the West Java Provincial Tourism and Culture Office (Dinas Pariwisata dan Kebudayaan Provinsi Jawa Barat, Disparbud Jabar). Assembling the corpus from multiple public secondary sources, rather than reusing a single existing dataset, reflects the absence of a publicly released tourism dataset that offers broad coverage of the three administrative units of Bandung Raya with the four attributes required by this study (name, description, category, and region).

Two filtering rules were applied to the collected records. First, destinations located outside Bandung City, Bandung Regency, and West Bandung Regency were excluded so that the corpus is aligned with the study's regional scope. Second, duplicate records and records without a sufficiently descriptive text field were removed. After these two rules were applied, the retained corpus contains 331 destinations. Because collection and filtering were interleaved rather than logged as a strict per-stage pipeline, the number of records at intermediate stages was not recorded; this documentation gap is acknowledged as a limitation.

The resulting corpus is heterogeneous in both category and region. Nature attractions dominate with 157 destinations (47.4%), followed by city parks with 36 (10.9%), shopping centres with 21 (6.3%), and museums with 10 (3.0%), while tourism villages form the smallest group with 3 destinations (0.9%); the full category distribution is shown in Fig. 2. Geographically, the corpus is well balanced across the three administrative areas: Bandung Regency contributes 126 destinations (38.1%), Bandung City 118 (35.6%), and West Bandung Regency 87 (26.3%). This imbalance across categories is important for interpreting the evaluation results, because a query targeting a sparsely populated category cannot retrieve ten relevant items even from a perfect ranking.

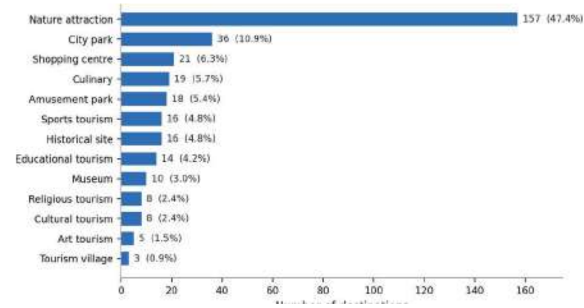

Fig.2. Category distribution of the destination corpus

3.2. Text Preprocessing

The preprocessing stage includes five processes, namely (1) case folding, converting all letters to lowercase; (2) cleaning, removing punctuation, numbers, and symbols; (3) tokenization, splitting text into words; (4) stopword removal, removing common words that carry no meaning; and (5) stemming, converting affixed words into their base form using the Sastrawi library for the Indonesian language.

3.3. TF-IDF Weighting

The preprocessing results are represented as numerical vectors using TF-IDF. The weight of each word is calculated through Equation (1) and Equation (2).

$$W(t,d) = TF(t,d) \times IDF(t) \tag{1}$$
$$IDF(t) = \ln [(1 + N) / (1 + DF(t))] + 1 \tag{2}$$

where W(t,d) is the weight of word t in document d, TF(t,d) the frequency of word t in document d, N the total number of documents, and DF(t) the number of documents containing word t. Weighting is performed using TfidfVectorizer from Scikit-learn and produces a 331 × 583 matrix.

Because the implementation uses scikit-learn's TfidfVectorizer with its default configuration, the inverse document frequency follows the smoothed form given in Equation (2) rather than the unsmoothed log(N/DF), and each document vector is subsequently normalised to unit L2 length. Only the preprocessed destination description is vectorised; the name, category, and region attributes are retained solely for filtering and display and do not enter the TF-IDF representation.

The vectoriser was applied with its default settings: tokens are lower-cased, only unigrams are used (ngram_range = (1, 1)), no term is filtered by document frequency (min_df = 1, max_df = 1.0), the vocabulary size is unbounded (max_features = None), term frequencies are linear (sublinear_tf = False), and both use_idf and smooth_idf are enabled. The comparatively small vocabulary of 583 terms is a direct consequence of the corpus itself: the destination descriptions are short (typically only a couple of sentences), and Sastrawi stemming together with Indonesian stopword removal further collapses morphological variants onto

Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.28

shared root forms. TF-IDF term weighting originates from classical information retrieval [32] and builds on the interpretation of inverse document frequency as a measure of term specificity [33]; in this study it is computed using the scikit-learn library [34].

3.4. Cosine Similarity Computation

The similarity between destinations and between a user query and destinations is computed using Cosine Similarity as in Equation (3), namely by measuring the cosine of the angle between two TF-IDF vectors.

$$\mathrm{sim}(A, B) = (\Sigma A_i \times B_i) / (\sqrt{(\Sigma A_i^2)} \times \sqrt{(\Sigma B_i^2)}) \quad (3)$$

where A and B are the TF-IDF vectors of two documents. This computation produces a 331 × 331 similarity matrix that becomes the basis for generating recommendations.

3.5. System Architecture

The system is built as a website with a client-server architecture. The backend uses the Flask framework (Python) to process data, perform computations, and provide REST API endpoints [16]. The frontend uses the React framework (JavaScript) to display the interface and communicate with the backend through the REST API in JSON format [17]. The TF-IDF model and similarity matrix are stored in a .pkl file so that the backend does not repeat the entire process each time it runs.

The deployed system exposes two distinct recommendation paths. For item-to-item recommendation, when a user opens a destination the corresponding row of the precomputed 331 × 331 similarity matrix is read directly and the highest-scoring destinations are returned, so no vectorisation is performed at request time. For free-text queries, the user's text is passed through the same five preprocessing stages and then projected into the existing vocabulary using the fitted vectoriser's transform operation (not a new fit), which guarantees that the query vector shares the 583-term space of the corpus; its cosine similarity against the 331 × 583 document matrix is then computed and the top-N destinations are returned. Using a single fitted vectoriser for both paths ensures vocabulary consistency between the corpus and user queries.

3.6. System Testing

Functional testing uses Black-Box Testing to validate each feature based on the correspondence between input and output [18]. Recommendation accuracy is measured using Precision@K, which calculates the proportion of relevant recommendations in the top K (Equation 4), and Mean Average Precision (MAP), namely the average of the Average Precision of all test queries (Equation 5). Rank-aware measures such as Precision@K and Mean Average Precision are standard tools for evaluating the quality of recommender-system output [35], and have been applied across a range of recommendation domains including research-paper recommendation [36].

$$\text{Precision@K} = (\text{relevant in Top-K}) / K \quad (4)$$
$$\text{MAP} = (1 / Q) \times \Sigma\, AP_q \quad (5)$$

where Q is the number of queries and AP_q the Average Precision of the q-th query. Testing uses five queries representing various categories, namely "museum", "hot spring", "town square", "waterfall", and "city park", with relevance determined manually using a binary scale.

Average Precision for a single query is defined as $AP = (1/R) \times \Sigma\, P(k)\cdot rel(k)$ summed over ranks k = 1 to K, where K = 10 is the cut-off, P(k) is the precision at rank k, $rel(k) \in \{0, 1\}$ indicates whether the item at rank k is relevant, and R is the number of relevant items for that query. The number of relevant items used was 10 for the museum, hot spring, waterfall, and city park queries, and 5 for the town square query, reflecting the fact that fewer than ten town-square destinations exist in the corpus. MAP is the mean of these per-query AP values.

The five queries — museum, hot spring, town square, waterfall, and city park — were chosen to span categories of differing frequency in the corpus, from densely populated categories to a sparse one, so that the evaluation probes retrieval behaviour across the category distribution rather than in a single favourable case. Relevance was judged by a single assessor (the author) against written binary criteria defined per query before the results were inspected. Because only one assessor was used, inter-annotator agreement (for example, Cohen's kappa) could not be computed; the evaluation over five queries with a single assessor is therefore best read as an initial, internally consistent assessment rather than a definitive benchmark. Extending the query set and introducing independent annotators are left to future work.

The named top-10 destinations retrieved for each of the five test queries, together with their assigned binary relevance labels, are listed in the Appendix (Table A1) for reproducibility. For the town square query in particular, the relevant items occupy ranks 1, 2, 3, 4, and 7, consistent with the worked calculation given above.

4. RESULTS AND DISCUSSIONS

4.1. System Implementation

The system was successfully implemented as a website with four main pages, namely Home, Destination List, Destination Detail, and About. The system provides two search modes, namely "Search by Place" to find destinations by name, and "Tell Us Your Wish" which allows users to enter a free description in Indonesian. On the Destination Detail page, the system displays six similar destinations based on Cosine Similarity. Separating the TF-IDF fitting and vectorisation stage from the recommendation-serving stage provides computational efficiency and vocabulary

BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS
Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.28

consistency between the fitted corpus and user queries. The deployed interface is illustrated in Fig. 3, which shows the detail page of a recommended destination, and Fig. 4, which shows the six most similar destinations listed beneath it, retrieved directly from the precomputed similarity matrix.

Fig.3. Detail page of a recommended destination

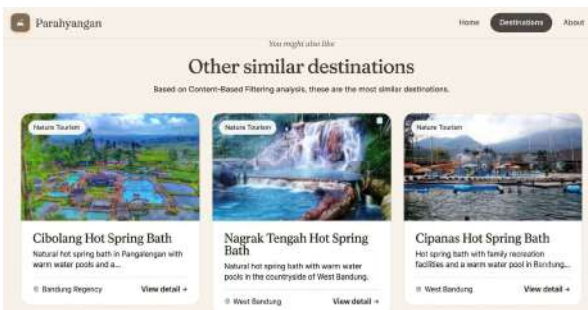

Fig.4. Six most similar destinations shown on the detail page

4.2. Functional Testing

Functional testing was conducted on eight scenarios. All scenarios ran successfully according to the expected results without any errors found, as summarized in Table 1.

TABLE 1. Black-Box Testing Results

Code	Test Scenario	Status
KF-01	Open the Home page	Passed
KF-02	Open the Destination List (331 destinations)	Passed
KF-03	Region filter on the Destination List	Passed
KF-04	Category filter on the Destination List	Passed
KF-05	Open destination detail and 6 similar destinations	Passed
KF-06	Search via "Search by Place"	Passed
KF-07	Free description via "Tell Us Your Wish"	Passed
KF-08	Open the About page	Passed

4.3. Accuracy Testing

Accuracy testing used five test queries. The Precision@5, Precision@10, and Average Precision (AP) values for each query are shown in Table 2, with an average Precision@5 of 96.00%, Precision@10 of 90.00%, and MAP of 98.86%.

TABLE 2. Precision@K and MAP Results

Query	P@5 (%)	P@10 (%)	AP (%)
museum	100.0	100.0	100.00
hot spring	100.0	100.0	100.00
town square	80.0	50.0	94.29
waterfall	100.0	100.0	100.00
city park	100.0	100.0	100.00
Average / MAP	96.00	90.00	98.86

MAP is calculated as the average of the Average Precision (AP) values of all test queries.

4.4. Qualitative Analysis

To illustrate the behaviour of the model beyond aggregate metrics, Table 3 shows the six destinations ranked most similar to Museum Geologi. The highest-scoring item is Museum Pendidikan Nasional, with a Cosine Similarity of 0.3181, and five of the six top-ranked destinations belong to the museum category, providing preliminary evidence that the TF-IDF representation captures the thematic content of the descriptions.

TABLE 3. Top-6 destinations most similar to Museum Geologi

Rank	Destination	Category	Cosine similarity
1	Museum Pendidikan Nasional	Museum	0.3181
2	Museum Pos Indonesia	Museum	0.2648
3	Museum Kota Bandung	Museum	0.2644
4	Museum Sri Baduga	Museum	0.2533
5	Museum Mandala Wangsit Siliwangi	Museum	0.2452
6	Stone Garden Citatah	Nature attraction	0.2246

The remaining item is more instructive. Stone Garden Citatah is labelled as a nature attraction, yet it is retrieved among the top six because its description overlaps lexically with that of the reference destination. This shows that the system ranks by textual similarity rather than by category label, which is precisely the intended behaviour of Content-Based Filtering: it can surface cross-category items that a simple category filter would never return. The same mechanism, however, explains the lower Precision@10 observed for narrow queries, since lexically similar but categorically different destinations may be judged irrelevant under a strict binary relevance scheme. The absolute similarity values are also modest in magnitude, which is expected for sparse TF-IDF vectors over short descriptions; what matters for recommendation is the relative ordering rather than the absolute score.

4.5. Discussion

The MAP of 98.86% indicates, for the five evaluated queries, that relevant destinations are consistently ranked at the top of the result list, which is the property that matters most in a recommendation setting where users inspect only the first few items. The gap between Precision@5 (96.00%) and Precision@10 (90.00%) is expected: as K grows beyond the number of genuinely relevant items in the corpus, precision necessarily degrades. This effect is clearest for the query "town square", whose Precision@10 falls to 50.0% simply because fewer than ten truly relevant destinations exist in the dataset, while its AP remains high at 94.29% because the relevant items still occupy the top ranks.

For completeness, the town-square case can be reconstructed from the reported figures. With five relevant

Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.28

items (R = 5) located at ranks 1, 2, 3, 4, and 7 of the top ten, Average Precision evaluates to (1/5) × (1/1 + 2/2 + 3/3 + 4/4 + 5/7) = 0.9429, while Precision@5 = 4/5 = 80.0% and Precision@10 = 5/10 = 50.0%, exactly the pattern reported in Table 2.

The high accuracy is also attributable to the structure of the destination descriptions, in which discriminative keywords appear early and recur naturally, yielding higher TF-IDF weights. Because the model operates on lexical similarity rather than category labels, semantically related but lexically distinct destinations may occasionally be retrieved; this is an inherent characteristic of the TF-IDF representation rather than an implementation error. Compared with prior work on Yogyakarta [9] and Bali [6], a direct numerical comparison with those studies is not meaningful, because the corpora, query sets, and relevance judgements differ; the present results are reported on their own terms rather than as evidence of superiority.

The main limitations are threefold. First, the system relies exclusively on Content-Based Filtering and therefore cannot exploit preference signals from other users, which restricts serendipity and tends to produce homogeneous recommendations. Second, the corpus is limited to 331 destinations. Third, ratings and popularity are not incorporated into the ranking function. In practical terms, however, the deployed system already provides a usable decision-support tool for tourists and a low-cost channel through which less prominent destinations can become discoverable, which is the primary applied benefit of the approach.

5. CONCLUSION

This study developed and deployed a machine learning-based recommender system for tourism destinations in Bandung Raya using Content-Based Filtering with TF-IDF and Cosine Similarity. A corpus of 331 destinations was preprocessed with the Sastrawi library, vectorised into a 331 × 583 TF-IDF matrix, and compared to yield a 331 × 331 similarity matrix. The model was served through a Flask, React, and REST API website offering name-based and free-text query modes. All eight Black-Box Testing scenarios passed, and the system achieved Precision@5 of 96.00%, Precision@10 of 90.00%, and MAP of 98.86%, providing preliminary evidence that relevant destinations are ranked highly for the five evaluated queries. Future work will extend the corpus, incorporate Collaborative Filtering into a hybrid architecture to mitigate the homogeneity of content-based recommendations, and explore contextual embeddings such as word2vec or transformer-based sentence encoders as an alternative to TF-IDF.

REFERENCES

[1] D. Sutrisno, “Angka Kunjungan Wisatawan ke Bandung Naik Signifikan pada 2024,” IDN Times Jabar, 2025. [Online]. Available: https://jabar.idntimes.com/news/jawa-barat/angka-kunjungan-wisatawan-ke-bandung-naik-signifikan-pada-2024-00-9y7yl-xrvkyg

[2] D. M. Saputra, N. Angelia, and N. Yusliani, “Recommender system for tourist destinations in Indonesia using matrix factorization method,” JITSI, vol. 5, no. 3, pp. 122–127, 2024, doi: 10.62527/jitsi.5.3.254.

[3] I. Setianingsih and S. Widiono, “Inovasi aplikasi smart tourism berbasis mobile untuk optimalisasi informasi destinasi wisata,” Edumatic: J. Pendidik. Inform., vol. 9, no. 3, pp. 806–814, 2025, doi: 10.29408/edumatic.v9i3.32607.

[4] D. Roy and M. Dutta, “A systematic review and research perspective on recommender systems,” J. Big Data, vol. 9, art. 59, 2022, doi: 10.1186/s40537-022-00592-5.

[5] F. Christyawan, A. N. Rohman, and A. D. Hartanto, “Application of content-based filtering method using cosine similarity in restaurant selection recommendation system,” J. Inf. Syst. Informatics, vol. 6, no. 3, pp. 1559–1576, 2024, doi: 10.51519/journalisi.v6i3.806.

[6] R. Faurina and E. Sitanggang, “Implementasi metode content-based filtering dan collaborative filtering pada sistem rekomendasi wisata di Bali,” Techno.Com, vol. 22, no. 4, pp. 870–881, 2023, doi: 10.33633/tc.v22i4.8556.

[7] B. Prasetyo, V. Atina, and E. Purwanto, “Sistem rekomendasi pariwisata dengan metode content based recommendation berbasis website,” DutaCom, vol. 14, no. 1, pp. 51–58, 2021, doi: 10.47701/dutacom.v14i1.2017.

[8] W. Yulita et al., “Automatic scoring using term frequency inverse document frequency and cosine similarity,” Sci. J. Informatics, vol. 10, no. 2, pp. 93–104, 2023, doi: 10.15294/sji.v10i2.42209.

[9] J. F. U. Albab and A. N. Rohman, “Recommendation system Yogyakarta tourism using TF-IDF and cosine similarity,” J. Appl. Informatics Comput., vol. 10, no. 1, pp. 935–945, 2026, doi: 10.30871/jaic.v10i1.11751.

[10] C. Salsabilla and D. W. Utomo, “Pekalongan Regency tourism recommendation system with content based filtering,” SATIN, vol. 14, no. 1, pp. 262–270, 2025, doi: 10.32520/stmsi.v14i1.4839.

[11] A. W. Wicaksono, A. N. Rohman, and A. D. Hartanto, “Café recommendation using the content-based filtering method,” J. Inf. Syst. Informatics, vol. 6, no. 3, pp. 1598–1615, 2024, doi: 10.51519/journalisi.v6i3.813.

[12] A. Y. Timur, “Sistem rekomendasi lagu Indonesia menggunakan metode content-based filtering dan cosine similarity,” J. Informatika Tek. Elektro Terap., vol. 13, no. 1, pp. 1415–1423, 2025, doi: 10.23960/jitet.v13i1.5949.

[13] G. H. Setiawan and I. M. B. Adnyana, “Improving helpdesk chatbot performance with TF-IDF and cosine similarity models,” J. Appl. Informatics Comput., vol. 7, no. 2, pp. 252–257, 2023, doi: 10.30871/jaic.v7i2.6527.

[14] Y. S. Pasaribu and T. S. Sitompul, “Rekomendasi destinasi wisata Kota Bandung menggunakan algoritma collaborative filtering,” J. Mutiara, vol. 1, no. 6, pp. 382–392, 2023, doi: 10.59059/mutiara.v1i6.736.

[15] Y. Anis, E. N. Wahyudi, and H. C. Kurniawan, “Metode Waterfall dalam pengembangan sistem inventaris guna meningkatkan efisiensi manajemen stok barang,” J. Teknol. dan Sist. Inf. Bisnis, vol. 6, no. 2, pp. 329–338, 2024, doi: 10.47233/jteksis.v6i2.1351.

[16] H. L. Walingkas and P. O. N. Saian, “Penerapan framework Flask pada pembangunan sistem informasi pemasok barang,” J. JTIK, vol. 7, no. 2, 2023, doi: 10.35870/jtik.v7i2.729.

[17] E. Naresvari and Y. A. Susetyo, “Penerapan JavaScript React pada perancangan front-end website UMKM Jemari Ragil,” IT-EXPLORE, vol. 4, no. 1, pp. 16–32, 2025, doi: 10.24246/itexplore.v4i1.2025.pp16-32.

[18] M. Zen, I. Irwan, H. Hafni, and M. D. P. Ananda, “Implementasi dan pengujian menggunakan metode black box testing pada sistem informasi tracer study,” Bull. Comput. Sci. Res., vol. 4, no. 4, pp. 327–340, 2024, doi: 10.47065/bulletincsr.v4i4.359.

[19] Government of the Republic of Indonesia, Peraturan Presiden Republik Indonesia No. 45 Tahun 2018 tentang Rencana Tata Ruang Kawasan Perkotaan Cekungan Bandung [Presidential Regulation No. 45 of 2018 on the Spatial Plan of the Bandung Basin Urban Area], State Gazette of the Republic of Indonesia No. 91, Jakarta, 2018.

[20] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, “Using collaborative filtering to weave an information tapestry,” Communications of the ACM, vol. 35, no. 12, pp. 61–70, 1992, doi: 10.1145/138859.138867.

Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.28

[21] P. Resnick and H. R. Varian, "Recommender systems," Communications of the ACM, vol. 40, no. 3, pp. 56–58, 1997, doi: 10.1145/245108.245121.
[22] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005, doi: 10.1109/TKDE.2005.99.
[23] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," Knowledge-Based Systems, vol. 46, pp. 109–132, 2013, doi: 10.1016/j.knosys.2013.03.012.
[24] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments: A survey," Decision Support Systems, vol. 74, pp. 12–32, 2015, doi: 10.1016/j.dss.2015.03.008.
[25] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," Egyptian Informatics Journal, vol. 16, no. 3, pp. 261–273, 2015, doi: 10.1016/j.eij.2015.06.005.
[26] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in The Adaptive Web, LNCS vol. 4321, Springer, 2007, pp. 325–341, doi: 10.1007/978-3-540-72079-9_10.
[27] P. Lops, M. de Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in Recommender Systems Handbook, Springer, 2011, pp. 73–105, doi: 10.1007/978-0-387-85820-3_3.
[28] J. Son and S. B. Kim, "Content-based filtering for recommendation systems using multiattribute networks," Expert Systems with Applications, vol. 89, pp. 404–412, 2017, doi: 10.1016/j.eswa.2017.08.008.
[29] R. Burke, "Hybrid recommender systems: Survey and experiments," User Modeling and User-Adapted Interaction, vol. 12, no. 4, pp. 331–370, 2002, doi: 10.1023/A:1021240730564.
[30] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," Advances in Artificial Intelligence, vol. 2009, art. 421425, 2009, doi: 10.1155/2009/421425.
[31] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in Recommender Systems Handbook, Springer, 2011, pp. 1–35, doi: 10.1007/978-0-387-85820-3_1.
[32] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Information Processing & Management, vol. 24, no. 5, pp. 513–523, 1988, doi: 10.1016/0306-4573(88)90021-0.
[33] K. Spärck Jones, "A statistical interpretation of term specificity and its application in retrieval," Journal of Documentation, vol. 28, no. 1, pp. 11–21, 1972, doi: 10.1108/eb026526.
[34] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[35] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," ACM Transactions on Information Systems, vol. 22, no. 1, pp. 5–53, 2004, doi: 10.1145/963770.963772.
[36] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: A literature survey," International Journal on Digital Libraries, vol. 17, no. 4, pp. 305–338, 2016, doi: 10.1007/s00799-015-0156-0.

APPENDIX

This appendix lists the named top-10 destinations retrieved for each of the five test queries used in the accuracy evaluation (Section 4.3), together with the cosine similarity and the binary relevance label.

TABLE A1. Top-10 destinations returned for each test query (1 = relevant, 0 = not relevant).

Query	Rank	Destination	Category	Region	Cosine sim.	Rel.
museum	1	Museum Pendidikan Nasional	Museum	Bandung City	0.3364	1
	2	Museum Asia Afrika	Museum	Bandung City	0.2855	1
	3	Museum Pos Indonesia	Museum	Bandung City	0.2786	1
	4	Museum Kota Bandung	Museum	Bandung City	0.2782	1
	5	Museum Sri Baduga	Museum	Bandung City	0.2664	1
	6	Gedung Merdeka	Historical site	Bandung City	0.2641	1
	7	Museum Mandala Wangsit Siliwangi	Museum	Bandung City	0.2580	1
	8	Gedung Indonesia Menggugat	Historical site	Bandung City	0.2546	1
	9	Museum Nike Ardilla	Museum	Bandung City	0.2508	1
	10	Museum NuArt Sculpture Park	Art tourism	Bandung City	0.2486	1
hot spring	1	Pemandian Air Panas Walini	Nature attraction	Bandung Regency	0.6468	1
	2	Pemandian Air Panas Cibolang	Nature attraction	Bandung Regency	0.6436	1
	3	Pemandian Air Panas Nagrak Tengah	Nature attraction	West Bandung Regency	0.6421	1
	4	Pemandian Air Panas Cipanas	Nature attraction	West Bandung Regency	0.6383	1
	5	Maribaya Hot Spring	Nature attraction	Bandung Regency	0.5331	1
	6	Cibolang Hot Spring	Nature attraction	Bandung City	0.5085	1
	7	Pemandian Air Panas Ciwidey	Nature attraction	Bandung Regency	0.4867	1
	8	Sari Ater Hot Spring	Nature attraction	West Bandung Regency	0.4704	1
	9	Ranca Upas	Nature attraction	Bandung Regency	0.4653	1
	10	Curug Cipanas	Nature attraction	West Bandung Regency	0.4355	1
town square	1	Alun-Alun Cicendo	City park	Bandung City	0.7185	1
	2	Alun-Alun Regol	City park	Bandung City	0.6590	1
	3	Alun-Alun Ujung Berung	City park	Bandung City	0.6017	1
	4	Alun-Alun Bandung	City park	Bandung City	0.5489	1
	5	Puncak Bintang	Nature attraction	Bandung Regency	0.0000	0
	6	Dago Pakar	Nature attraction	Bandung Regency	0.0000	0
	7	Taman Hutan Raya Djuanda	Nature attraction	Bandung Regency	0.0000	1
	8	Goa Jepang	Historical site	Bandung Regency	0.0000	0
	9	Goa Belanda	Historical site	Bandung Regency	0.0000	0
	10	Situ Patengan	Nature attraction	Bandung Regency	0.0000	0
waterfall	1	Curug Bentang Padjadjaran	Nature attraction	Bandung Regency	0.3690	1

Vol. 1 No.5 July 2026
E-ISSN: 3132-5115
DOI:10.65780/bima.v1i5.28

Query	Rank	Destination	Category	Region	Cosine sim.	Rel.
	2	Curug Sawer Cililin	Nature attraction	West Bandung Regency	0.3532	1
	3	Curug Aseupan	Nature attraction	West Bandung Regency	0.3443	1
	4	Wisata Sungai Cikapundung Maribaya	Nature attraction	Bandung Regency	0.3261	1
	5	Curug Halimun	Nature attraction	Bandung Regency	0.3260	1
	6	Curug Cipanji	Nature attraction	Bandung Regency	0.3233	1
	7	Curug Omas	Nature attraction	Bandung Regency	0.3222	1
	8	Curug Layung & Camping Ground	Nature attraction	West Bandung Regency	0.3186	1
	9	Curug Siliwangi	Nature attraction	Bandung Regency	0.3161	1
	10	Curug Cipanas	Nature attraction	West Bandung Regency	0.3070	1
city park	1	Taman Merdeka	City park	Bandung City	0.5017	1
	2	Taman Vanda	City park	Bandung City	0.4192	1
	3	Taman Balai Kota	City park	Bandung City	0.4165	1
	4	Taman Saturnus	City park	Bandung City	0.3977	1
	5	Taman Cibeunying	City park	Bandung City	0.3656	1
	6	Taman Pramuka	City park	Bandung City	0.3455	1
	7	Taman Cikapayang Dago	City park	Bandung City	0.3416	1
	8	Peta Park	City park	Bandung City	0.3406	1
	9	Taman Gedebage	City park	Bandung City	0.3395	1
	10	Kiara Artha Park	City park	Bandung City	0.3191	1

Notes: relevance was judged by a single assessor (the author) against the criteria defined in Section 3.6.