Journal of Applied Informatics Research Vol 2 No 2 (2026) pp. 13~19

Journal of Applied Informatics Research
Fakultas Vokasi, Kampus Unesa 1, Ketintang, Surabaya, Indonesia
Website: https://journal.unesa.ac.id/index.php/jair/ | E-mail: jair@unesa.ac.id

Development of a Content-Based Filtering Recommendation System for Indonesian Batik Using Motif Features and Regional Origins

Amallia Berliany Putri[1], Dodik Arwin Dermawan[2], I Gde Agung Sri Sidhimantra[3], Rosita[4]

[1,2,3,4]Manajemen Informatika, Universitas Negeri Surabaya, Indonesia
[1]amallia.22102@mhs.unesa.ac.id, [2]dodikdermawan@unesa.ac.id, [3]igdesidhimantra@mhs.unesa.ac.id, [4]rositarosita@unesa.ac.id

ARTICLE INFORMATION

Article history:

Received October 27, 2026
Revised November 04, 2026
Accepted November 09, 2026

Keywords:

Batik;
Recommender Systems;
Content-Based Filtering;
TF-IDF;
Cosine Similarity

ABSTRACT

Batik motifs vary in regional origin and cultural meaning, while relevant product information is scattered across sellers. This study developed BatikMatch, a web-based system that ranks batik motifs according to textual similarity. The dataset contained 75 motifs documented from four batik shops or MSMEs in East Java. Seven descriptive attributes (motif name, category, regional origin, color, production technique, philosophical meaning, and intended use) provided the basis for textual descriptions processed through case folding, tokenization, stopword removal, and stemming. TF-IDF weighting generated a 75 x 172 feature matrix, and cosine similarity yielded a 75 x 75 matrix. Using Sparkling as the reference motif, the four highest-ranked alternatives were Remo Surabayan (0.5430), Skena Surabaya (0.3815), Tugu Alun-Alun Mojokerto (0.3605), and Belimbing Lining Limo (0.3526). Reported black-box tests supported the observed interface and recommendation-display functions. Compared with wedding-oriented recommendation research, BatikMatch emphasizes general motif exploration using field-collected descriptions. Similarity scores indicate the degree of textual-vector resemblance, not recommendation accuracy or user satisfaction; independent relevance evaluation remains necessary.

1. INTRODUCTION

Batik is an important element of Indonesian cultural heritage, embodying aesthetic values, local identity, and philosophical meaning through its diverse patterns. Batik motifs vary in their regional origins, visual forms, colors, and symbolic meanings [1], [2]. Previous studies have examined UNESCO's recognition of batik [3], its preservation among younger generations [4], [5], and the geometric characteristics of batik motifs [6]. The wide range of motifs and the fragmented presentation of product information can make it difficult for users to select batik that matches specific needs or preferences. Digital technologies provide opportunities both to disseminate cultural information and to help users explore motif diversity more systematically [7], [8].

Recommender systems help users narrow down choices by analyzing item characteristics or patterns of user interaction. One approach is content-based filtering (CBF), in which recommendations are generated by comparing item attributes or content representations [9], [10]. Alternative CBF models using singular value decomposition have also been investigated [11]. For batik, attributes such as motif category, regional origin, color, production technique, philosophical meaning, and intended style of use can be represented as text and compared to identify similarities between motifs. This approach allows the system to generate recommendations without requiring historical ratings from a large user population.

Previous research has addressed batik recommendation with different decision contexts. Al Aqthar and Prapanca [12] used content-based filtering and cosine similarity to match motifs with traditional wedding contexts;

the study described approximately 29 motifs, incorporated motif-related and occasion information, and reported an expert assessment averaging 3.3 on a five-point scale. Khotimah et al. [13] addressed batik-product sales with collaborative filtering, Saputro et al. [14] used knowledge-based selection of batik clothing, and Ihsan and Suharjo [15] employed the certainty-factor method. Thus, neither content-based filtering nor cosine similarity is new to batik recommendation. Table 2 positions BatikMatch against these related approaches.

A remaining application gap is a motif-exploration system that organizes descriptions collected from multiple batik sellers and presents similarity-ranked alternatives without relying on wedding-specific requirements or historical user ratings. BatikMatch addresses this context by documenting 75 motifs from four East Java sources and representing motif characteristics through seven descriptive fields. The text is processed with TF-IDF and ranked using cosine similarity. This is a contribution in dataset organization, attribute integration, and practical application scope, not a claim of a novel ranking algorithm.

The objectives are to (1) develop the BatikMatch prototype, (2) demonstrate the text-based representation and ranking pipeline on 75 motif records, and (3) document its reported functional behavior. The resulting contribution is a reproducible application design and a descriptive case study of motif similarity, subject to verification of the text-processing configuration. Because the reported experiments contain only functional tests, they do not establish recommendation relevance, ranking accuracy, or user satisfaction; those claims require independent evaluation.

Table 1. Comparison of related batik recommendation research

Study	Approach	Reported application	Distinction from BatikMatch
Al Aqthar and Prapanca [12]	CBF + cosine similarity	Wedding stages, roles and motif suitability; about 29 motifs; expert assessment (mean 3.3/5)	BatikMatch instead supports general motif browsing across 75 shop-collected records; no expert relevance evaluation yet
Khotimah et al. [13]	Collaborative filtering	Batik-cloth and clothing sales recommendation	Item-metadata similarity rather than dependence on collaborative history
Saputro et al. [14]	Knowledge-based	Selection of batik clothing	Text-vector ranking of recorded motif features
Ihsan and Suharjo [15]	Certainty factor	Recommendation of batik clothing	Similarity-based ranking rather than certainty-factor rules

2. RESEARCH METHODS

2.1. Research Design and Data Collection

The study followed a research-and-development (R&D) approach using the prototyping model, comprising requirements communication, quick planning, initial design, development, and testing. Motif records were gathered through interviews, observation, and direct documentation at batik MSMEs and shops in East Java. Each record was organized into named descriptive fields for processing. The documented materials establish the sources and counts, but do not describe a formal motif-inclusion rule, duplicate-resolution procedure, or independent validation of every recorded attribute. Such details should be reported after verification against the original field records.

The final dataset comprised 75 motifs collected from four sources. Table 2 presents the locations and numbers of records obtained. These sources indicate the shops from which motif data were collected; the geographic origin of an individual motif was stored as a separate attribute and does not necessarily correspond to the location of the shop.

Table 2. Sources and numbers of batik motif records

No	Data source	Shop location	Count
1	Batik Nusantara	Surabaya	12
2	Tita Collection Batik	Bojonegoro	12
3	Rumah Batik Suramadu	Surabaya	17
4	Batik Bu Dar	Mojokerto	34
Total			75

2.2. Representation of Batik Motif Attributes

Each motif record contains seven candidate descriptive attributes (Table 4). Textual representation permits a content-based model to compare categorical labels (such as region or production technique) together with longer linguistic fields (such as philosophical meaning) within a shared term vocabulary. The manuscript describes concatenating the recorded fields into one motif description; price and images are available in the application but

were not used as textual features. No explicit per-attribute multiplier is documented. This does not mean all attributes have equal influence: longer fields may contribute more tokens, and TF-IDF assigns higher weight to terms that are rarer in the corpus.

Table 3. Attributes used to construct motif descriptions

Attribute	Information represented
Motif name	Name or identifying designation of a batik motif
Category	Pattern type, such as figurative, floral, faunal, geometric, or nature-inspired
Region of origin	Geographic origin associated with the motif
Color	Dominant color of the motif or fabric
Production technique	Production method, such as hand-drawn or stamped batik
Philosophical meaning	Cultural values or symbolic meanings associated with the motif
Intended use	Usage context, such as formal or casual wear

For example, the recorded description for Sparkling is "motif figuratif khas surabaya warna oren teknik cap melambangkan kemegahan dengan gaya formal" (a figurative Surabaya motif, orange in color, stamped, symbolizing grandeur, for formal wear). Kembang Bungur is described as "motif flora khas surabaya warna kuning teknik cap melambangkan keindahan dengan gaya kasual" (a floral Surabaya motif, yellow in color, stamped, symbolizing beauty, for casual wear). These Indonesian strings are retained because they are examples of the text input. They also show that the displayed excerpts contain category, location, color, technique, meaning, and use, but not the motif-name token itself; whether the name was included in the actual vectorizer input should be confirmed from the source data and code.

2.3. Text Preprocessing and TF-IDF Weighting

Each recorded description underwent case folding, tokenization, stopword removal, and stemming before TF-IDF weighting [16]. Let f(t,d) denote the occurrence count of term t in motif description d, df(t) the number of descriptions containing t, and N the number of descriptions (N = 75). A conventional illustrative definition is TF(t,d) = f(t,d) and IDF(t) = ln[N/df(t)], with TF-IDF calculated as in Equation (1). This unsmoothed expression explains the principle of weighting but is not evidence of the exact settings used in the software. The implementation-specific TF variant, IDF smoothing, stopword source or list, Indonesian stemmer and its version, tokenization rules, vocabulary filtering, and vector-normalization settings must be checked against the preprocessing code to reproduce the reported numerical matrix.

$$TF\text{-}IDF(t, d) = TF(t, d) \times IDF(t) \tag{1}$$

In Equation (1), t denotes a term and d a motif description. TF-IDF assigns each motif a vector over the shared vocabulary, and terms occurring in both compared descriptions can contribute to their dot product. The 75 x 172 matrix reported in Section 3.1 contains 75 motif records and 172 term features. Exact feature weights depend on the implementation parameters, not solely on the general formula.

2.4. Cosine Similarity and Recommendation Ranking

Cosine similarity compares two TF-IDF vectors by the ratio of their dot product to the product of their Euclidean norms (Equation (2)) [9], [16]. Here A and B represent vectors of two motifs. For nonnegative TF-IDF vectors, scores range from 0 (no shared weighted direction) to 1 (identical vector direction). Shared tokens contribute to the numerator; the magnitude of that contribution depends on their TF-IDF weights and vector norms.

$$sim(A, B) = (A \cdot B) / (||A|| \times ||B||) \tag{2}$$

When a user selects a reference motif, the system compares its vector with the other 74 motifs, sorts the resulting scores in descending order, excludes self-similarity, and displays the top four alternatives. Percentages shown by the interface are obtained by multiplying cosine similarity by 100 (e.g., 0.5430 becomes 54.30%). They express relative textual similarity within this dataset, not predictive accuracy, relevance precision, or user satisfaction.

2.5. System Implementation and Testing

BatikMatch was developed as a web application with a frontend built using HTML, CSS, JavaScript, and Bootstrap [17]. The backend was implemented in Python with the Flask framework, while motif, MSME, and

administrator records were managed using MySQL [18]. Pandas and scikit-learn were also employed during development. The system architecture illustrates how user requests pass through the backend for recommendation processing and database access, as shown in Figure 1. Flask is one framework commonly used for Python-based web applications [19].

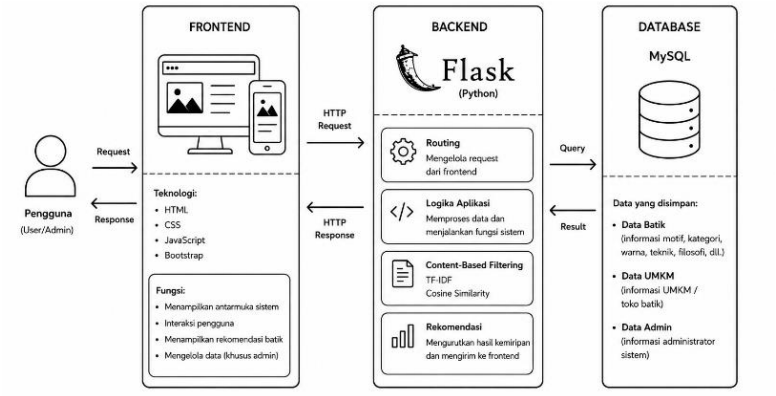


Figure 1. BatikMatch system architecture

Black-box testing assessed interface outputs for the home, catalog, recommendation, information, and administrative pages [20]. The available records document navigation controls, motif displays, and top-four recommendation presentation. The tests provide evidence of observed feature behavior rather than independent relevance judgments. Outcomes for malformed search inputs, invalid upload files, upload failures, and persisted add/edit/delete operations were not reported and therefore cannot be marked as passed.

3. RESULTS AND DISCUSSION

3.1. Dataset Processing Results

Data collection yielded 75 batik motif records from four sources, with Batik Bu Dar in Mojokerto contributing the largest number (34 motifs). The two sources in Surabaya contributed 12 and 17 motifs, respectively, while Tita Collection Batik in Bojonegoro supplied 12 motifs. The collected attributes were normalized into one textual description per motif. The dataset includes motifs associated with Surabaya, Bojonegoro, Madura, and Mojokerto, as illustrated by the examples in the research report.

After preprocessing, the 75 motif descriptions were transformed into a TF-IDF matrix with dimensions of 75 x 172. The first dimension represents the number of motif documents, whereas the second indicates the number of term features generated. In the weighting results reported for a subset of terms, the token "abhiboyo" in the third record had a TF-IDF weight of 0.319388, while the token "alam" (meaning "nature") in the fourth record had a weight of 0.389448. These are examples of weights in the numerical feature representation, not measures of overall system performance.

3.2. Pairwise Motif Similarity Results

The 75 motif vectors generated a 75 x 75 cosine similarity matrix. Each entry S(i,j) describes pairwise similarity between motifs i and j; diagonal entries are 1 for nonzero vectors compared with themselves. Table 5 lists four illustrative pairwise scores relative to Sparkling, whereas Table 6 presents the four highest-ranked recommendations. These two tables represent different subsets of the same ranking process.

Table 4. Example similarity scores relative to the Sparkling motif

Compared motif	Cosine similarity
Motif Kembang Bungur	0.233661
Motif Abhiboyo	0.306495
Motif Kintir-kintiran	0.200456
Motif Gembili Wonokromo	0.310016

The Sparkling and Kembang Bungur example descriptions share words associated with Surabaya and stamping, so those shared tokens can contribute positively to their cosine score (0.233661 in Table 5). Terms that differ, such as color, motif category, and style of use, do not contribute through identical-token overlap. Nevertheless, an exact numerical decomposition of 0.233661 or of the top-ranked score requires the two actual TF-IDF vectors, the vocabulary, and the normalization settings; these are not reported in the available results.

3.3. Four Highest-Ranked Recommendations

With Sparkling as the reference, the system excluded the reference record and ranked the remaining 74 motifs. Table 6 reports the four highest-scoring items, and Figure 2 shows the corresponding interface output.

Table 5. Four highest-ranked motifs recommended for Sparkling

No	Recommended motif	Cosine similarity	Similarity
1	Motif Remo Surabayan	0.5430	54.30%
2	Motif Skena Surabaya	0.3815	38.15%
3	Motif Tugu Alun-Alun Mojokerto	0.3605	36.05%
4	Motif Belimbing Lining Limo	0.3526	35.26%

The top score was 0.5430 for Remo Surabayan, 0.1615 higher than the second-ranked Skena Surabaya (0.3815). Tugu Alun-Alun Mojokerto (0.3605) and Belimbing Lining Limo (0.3526) followed closely. These gaps show how the algorithm orders motifs in its textual feature space; they do not show whether the ranked motifs are appropriate for a particular user or cultural context. Such judgments require an external relevance criterion rather than cosine similarity alone.

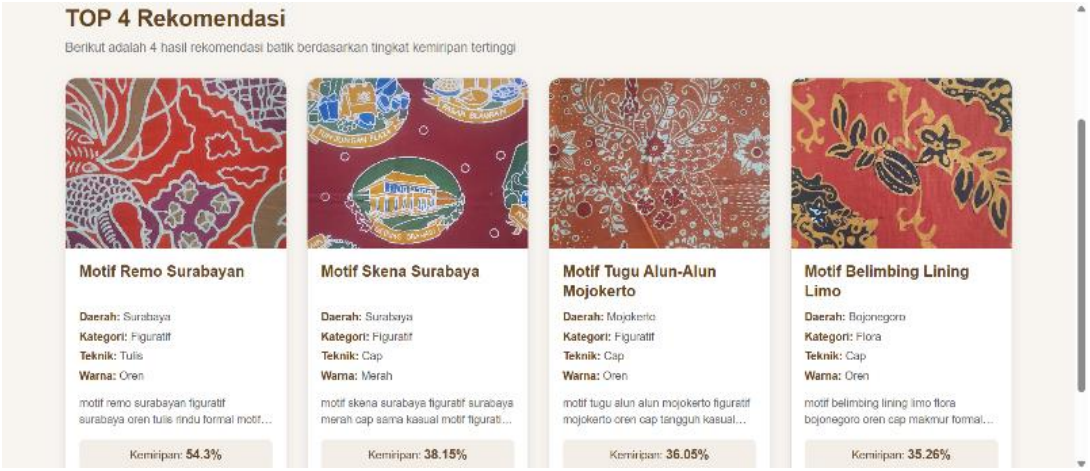


Figure 2. Four BatikMatch recommendations for the Sparkling reference motif (source: undergraduate final project report; original interface language retained)

3.4. Functional Testing Results

The final-project testing record documents navigation, catalog presentation, recommendation display, informational pages, dashboard summaries, and dataset-upload controls. Table 6 distinguishes the expected interface behavior from the observations reported for these scenarios. The status column refers only to the scope actually checked; visible controls do not demonstrate successful database updates or error handling. No additional tests were conducted for this revision.

Table 6. Summary of BatikMatch functional testing results

Test area	Scenario	Observed outcome
Home page	Display navigation and access to recommendations	Navigation controls and recommendation button were displayed.
Motif catalog	Search and display motif details	Search, motif cards, and detail page were displayed.
Recommendations	Select a motif and display the top four results	Four results appeared, ranked by similarity score.
Information pages	Display system information and contact details	System details, contact information, and FAQs were displayed.
Admin dashboard	Display batik and MSME summaries	Counts, status information, and method summaries were displayed.
Dataset management	Display the upload page and file-selection controls	File-selection controls and an upload button were displayed.

As summarized in Table 6, the reported scenarios provide functional support for interface presentation and the display of four similarity-ranked motifs. They do not establish successful handling of corrupt or unsupported dataset files, invalid queries, or persisted record changes. These negative and data-integrity scenarios should be executed and documented before any claim of comprehensive black-box verification is made.

3.5. Discussion

The principal contribution of BatikMatch is its application of a seven-field motif-description workflow to a 75-record, multi-seller catalog, enabling general-purpose similarity-based motif browsing. This is distinct in application scope and data provenance from the wedding-specific recommendation task in [12]. However, [12] also applied content-based similarity and reported expert ratings (mean 3.3/5), whereas BatikMatch has not yet reported independent relevance judgments. Consequently, this study cannot claim a performance or validation advantage over [12]. Comparisons with collaborative filtering [13] and knowledge- or rule-based methods [14], [15] describe different recommendation paradigms, not empirical performance comparisons.

The Sparkling case illustrates the interaction of textual overlap, corpus-dependent inverse-document frequency, and cosine normalization. A high score indicates that two weighted descriptions point in similar directions, not that their photographs look alike or that an expert would choose one in place of the other. Combining fields into a single text introduces an important modeling trade-off: it accommodates descriptive metadata without specialized field encoders, but may implicitly favor longer text fields or infrequent tokens. No controlled field-weighting or ablation study was reported, so the influence of each attribute on the final ranking remains unquantified.

The findings should be interpreted within three evidential boundaries. First, the 75 records originate from four East Java sellers and do not represent all Indonesian batik. Second, observed black-box behavior verifies only the tested interface scenarios; the study lacks expert- or user-labeled relevance, Precision@K, Recall@K, or NDCG@K. Third, exact preprocessing and weighting settings, including possible field weighting, must be recovered from the code for full reproducibility. A rigorous next evaluation would specify several reference motifs, collect independent graded or binary relevance judgments for candidate motifs, and compute Precision@4 and/or NDCG@4 from those judgments. Expert agreement, attribute sensitivity, and comparison with an alternative model would further strengthen the evidence; no such results are claimed here.

4. CONCLUSION

This study developed BatikMatch, a web-based content-filtering prototype for comparing batik motif descriptions drawn from 75 records and seven candidate attributes. Its reported processing generated a 75 x 172 TF-IDF matrix and a 75 x 75 pairwise similarity matrix. For the Sparkling reference motif, the highest-ranked alternative was Remo Surabayan (cosine similarity 0.5430), followed by Skena Surabaya, Tugu Alun-Alun Mojokerto, and Belimbing Lining Limo. Documented black-box cases supported the observed interface and top-four ranking displays. The contribution is a general motif-exploration workflow and its field-collected catalog, not a new cosine-similarity algorithm or a validated improvement in recommendation quality. Independent relevance testing, explicit preprocessing settings, and stronger administrative/error-path tests are required before claims about recommendation effectiveness can be made.

ACKNOWLEDGEMENTS

The authors express their appreciation to Batik Nusantara, Tita Collection Batik, Rumah Batik Suramadu, and Batik Bu Dar for their cooperation during the collection and documentation of batik motif information. Their participation contributed to the compilation of the dataset used in developing BatikMatch.

REFERENCES

[1] A. A. Trixie, "Filosofi motif batik sebagai identitas bangsa Indonesia," Folio, vol. 1, pp. 1–9, 2020. https://doi.org/10.37715/folio.v1i1.1380

[2] E. F. Ningsih dan Misgiya, "Penerapan motif flora dan fauna sebagai inspirasi penciptaan karya batik tulis," Jurnal Ilmu Pengetahuan, Seni, dan Teknologi, vol. 2, no. 2, pp. 57–71, 2025. https://doi.org/10.62383/imajinasi.v2i2.619

[3] Y. N. Evita, A. Trihartono, and A. Prabhawati, "Pengakuan UNESCO atas batik sebagai warisan budaya tak benda (WBTB)," Majalah Ilmiah Dian Ilmu, vol. 21, no. 2, pp. 113–128, 2022, https://doi.org/10.37849/midi.v21i2.260

[4] M. M. Aliyah and N. Hasaniyah, "Peran generasi muda dalam melestarikan budaya bangsa melalui kerajinan batik," Maliki Interdisciplinary Journal, vol. 3, no. 2, pp. 347–355, 2025.

[5] M. S. Sono and M. P. Elisabeth, "Upaya meningkatkan kecintaan terhadap batik Nusantara pada generasi muda (pendekatan teori determinasi diri)," Prosiding Seminar Nasional Industri Kerajinan and Batik, vol. 5, no. 1 (edisi 2023), pp. D.02 1–9, published in 2024.

[6] D. N. Sulisawati, Y. Saputra, and N. Hidayatin, "Eksplorasi konsep rotasi geometris pada motif batik Labako Jember," Differential: Journal on Mathematics Education, vol. 2, no. 2, pp. 233–242, 2025, https://doi.org/10.32502/differential.v2i2.424

[7] N. M. Mitayani dan D. A. S. W. Astuti, "Pengembangan batik khas kebudayaan Indonesia dengan pengaruh teknologi," Prosiding Pekan Ilmiah Pelajar (PILAR), vol. 4, pp. 60–64, 2024.

[8] N. Fitria and Supriono, "Dampak teknologi and komunikasi terhadap pelestarian and budaya lokal," At-Tadabbur: Jurnal Penelitian Sosial Keagamaan, vol. 14, no. 2, pp. 123–136, 2024.

[9] I. Saifudin dan T. Widiyaningtyas, "Systematic literature review on recommender system: Approach, problem,

evaluation techniques, datasets," IEEE Access, vol. 12, pp. 19827–19847, 2024, https://doi.org/10.1109/ACCESS.2024.3359274

[10] K. Benabbes, K. Housni, K. El Mezouary, dan A. Zellou, "Recommendation system issues, approaches and challenges based on user reviews," Journal of Web Engineering, vol. 21, no. 4, pp. 1017–1054, 2022, https://doi.org/10.13052/jwe1540-9589.2143

[11] F. Colace, D. Conte, M. De Santo, M. Lombardi, D. Santaniello, and C. Valentino, "A content-based recommendation approach based on singular value decomposition," Connection Science, vol. 34, no. 1, pp. 2158–2176, 2022, https://doi.org/10.1080/09540091.2022.2106943

[12] A. Z. Al Aqthar dan A. Prapanca, "Sistem rekomendasi motif batik sesuai dengan kebutuhan acara pernikahan pengguna menggunakan metode Content-Based Filtering," Journal of Informatics and Computer Science (JINACS), vol. 6, no. 3, pp. 823–829, 2025, https://doi.org/10.26740/jinacs.v6n03.p823-829

[13] A. H. Khotimah, S. Sumarlinda, dan F. E. Nastiti, "Sistem rekomendasi penjualan kain dan baju batik di Toko Batik Tresno Dharma Karanganyar menggunakan metode Collaborative Filtering," Smart Comp, vol. 13, no. 4, pp. 853–862, 2024, https://doi.org/10.30591/smartcomp.v13i4.6976

[14] N. Saputro, V. Atina, dan D. Hartanti, "Sistem rekomendasi pemilihan jenis baju batik menggunakan metode Knowledge-Based di Batik Amarta," Jurnal Fasilkom, vol. 14, no. 2, pp. 451–457, 2024. https://doi.org/10.37859/jf.v14i2.7483

[15] F. M. Ihsan dan I. Suharjo, "Sistem rekomendasi baju batik dengan metode Certainty Factor (studi kasus: Ata Branded)," Kumpulan Jurnal Ilmu Komputer (KLIK), vol. 11, no. 2, pp. 140–152, 2024. https://doi.org/10.20527/klik.v11i2.685

[16] D. Meidelfi, Yulherniwati, I. Rahmayuni, T. Hidayat, dan D. Chandra, "TF-IDF implementation for similarity checker on the final project title," International Journal of Advanced Science Computing and Engineering, vol. 3, no. 1, pp. 40–52, 2021, https://doi.org/10.62527/ijasce.3.1.3

[17] M. F. Santoso, "Perbandingan efektivitas Bootstrap and Tailwind CSS dalam pengembangan UI web responsif," Jurnal Teknologi and Sistem Informasi Bisnis, vol. 7, no. 4, pp. 489–497, 2025, https://doi.org/10.47233/jteksis.v7i4.2260

[18] F. D. Silalahi, Manajemen Database MySQL (Structured Query Language). Semarang: Yayasan Prima Agus Teknik, 2022.

[19] R. Y. Azhari, "Web service framework: Flask dan FastAPI," Technology and Informatics Insight Journal, vol. 1, no. 1, pp. 80–87, 2022. https://doi.org/10.32639/tiij.v1i1.54

[20] Z. M. Alshamaa dan N. N. Saleem, "Black box software testing techniques: A literature review," Passer Journal of Basic and Applied Sciences, vol. 7, no. 2, pp. 1001–1011, 2025, https://doi.org/10.24271/psr.2025.508216.1963

AUTHOR BIOGRAPHY

	Amallia Berliany Putri is a student of the Informatics Management Study Program, Faculty of Vocational Studies, Surabaya State University.
	Dodik Arwin Dermawan is a lecturer at the Department of Informatics Management, Surabaya State University, Indonesia.
	I Gde Agung Sri Sidhimantra is a lecturer at the Department of Informatics Management, Surabaya State University, Indonesia.
	Rosita is a lecturer at the Department of Informatics Management, Surabaya State University, Indonesia.