Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.4, August 2025, pp. 1858~1867
e-ISSN: 2548-6861

News Recommendation System Using Content-Based Filtering through RSS Customization Service

Ida Ayu Widya Nandita [1]*, I Made Agus Dwi Suarjaya [2]*, I Putu Agung Bayupati [3]*
* Teknologi Informasi, Universitas Udayana
ayuwidyanandita@student.unud.ac.id [1], agussuarjaya@it.unud.ac.id [2], bayupati@unud.ac.id [3]

Article Info

Article history:

Received 2025-06-14
Revised 2025-07-27
Accepted 2025-07-30

Keyword:

Content-Based Filtering,
News Recommendation System,
RSS Feed.

ABSTRACT

News refers to stories or information about current events or incidents. Several news websites offer a service called RSS (Really Simple Syndication), which enables users to easily receive updates on the latest news. News RSS feeds are typically generated based on the order of publication time or general categories. The content of these news RSS feeds can be customized to align with user interests or preferences. A recommendation system can be utilized as an approach to customize RSS feeds. This study was conducted to design a system capable of generating RSS feeds based on news recommendations using the content-based TF-IDF method and cosine similarity. Data scraping and preprocessing of news articles from various RSS feeds of Indonesian news websites were automated using cron jobs. Content-based filtering modeling was carried out using TF-IDF and cosine similarity. The design and customization of RSS feeds were implemented in a Flask application and packaged within several endpoints. The recommendations generated based on user click interactions were reasonably relevant, as they successfully presented news titles similar to the clicked articles, with cosine similarity scores ranging from 0.2 to 1.0. The majority of respondents agreed that the recommended news articles were relevant to the articles they had clicked and aligned with their interests. The RSS feed evaluation yielded highly satisfactory results, with all aspects assessed in the user acceptance survey achieving an average score exceeding 80%, and the overall results of the customer satisfaction survey indicated scores starting from 90%.

I. Introduction

The advancement of the internet and the widespread use of electronic devices have significantly increased individuals' access to information. News refers to stories or information about current events or incidents [1]. News is a prominent source of information available on both traditional and digital platforms. Numerous national and international news websites are easily accessible to the public.

Several news sites provide a service known as RSS (Really Simple Syndication), which allows users to receive the latest news updates conveniently. According to [2], RSS feeds are still widely used by more than 30 million global websites and more than 200 thousand national websites to syndicate content to users who subscribe. RSS facilitates the syndication of either entire or partial content [3]. It enables users to view summaries of content from news websites without having to visit the sites directly. Typically, RSS feeds are formatted in XML files and accessed through RSS feed readers.

News RSS feeds are generally generated based on chronological publication order or broad categories without considering individual user interests or preferences. However, the content of these RSS feeds can be customized to better align with user interests. A recommendation system can serve as a practical approach to enable this customization.

This study adopts a content-based filtering approach to design a news recommendation system, which groups articles similar to those previously clicked by the user. A study done by [4] used Google News users' history data to predict users' news category interest using a Bayesian framework, then combined with an existing collaborative filtering mechanism to generate personalized news recommendations.

Experiments on the live traffic of the Google News website demonstrated that the hybrid method improves the quality of news recommendations and increases traffic to the site. Research conducted by [5] evaluated various text-weighting methods using Natural Language Processing (NLP), specifically Term Frequency-Inverse Document Frequency (TF-IDF), Latent Semantic Indexing (LSI), Doc2Vec, multilingual BERT (mBERT), and XLM-ROBERTa (XLM-R), to generate related news article recommendations. The findings showed that the TF-IDF method produced relatively consistent recommendation results with a mean average precision (MAP) for top 10 news recommendations (MAP@10) value of 0,281 compared to curated related-article lists created by journalists. A similar study done in [6] compared TF-IDF and Word2Vec embeddings for text's words and semantic information and calculated the similarity with cosine similarity and soft cosine similarity to create a list of news recommendations. The top-1 accuracy was calculated by deciding if the top-1 recommended news title's category was the same as that of the related news. The results show TF-IDF and cosine similarity achieved the highest top-1 accuracy of 76,8%. In another study, [7] implemented TF-IDF and cosine similarity to generate news recommendations based on users' news title click histories, achieving a hit rate of 80.77%, defined as the ratio between recommended and previously clicked news items. Meanwhile, [3] developed personalized RSS feed recommendations based on user profile categories defined through keyword ontology. However, their system was limited to recommending existing RSS feeds without performing direct RSS content customization.

These studies demonstrate the potential of TF-IDF combined with cosine similarity in generating relevant news recommendations based on related content and user behavior. In the context of this research, the method was also selected due to the characteristics of the dataset and the exploratory nature of the study. Specifically, the dataset lacks detailed user interaction history (such as reactions, likes, or dwell time) that are typically required to implement collaborative filtering or deep learning-based recommendation models. Given that the service is still in its early development phase, and sufficient user interaction data has not yet been collected, simpler content-based methods remain the most practical choice. The focus is to assess feasibility and user acceptance rather than benchmark against large-scale datasets.

This study aims to design a service-oriented system capable of generating customized RSS feeds based on news recommendation results that are aligned with user interests. Rather than merely providing generic RSS feeds, the proposed system customizes the feed content dynamically for each user. The system applies a content-based filtering method using TF-IDF and cosine similarity to generate recommendations. The resulting recommendation system is then implemented into a collection of news RSS feeds to produce outputs tailored to individual user preferences based on click activity.

II. Methods

This section discusses the methods used in this research to achieve the stated objectives. Figure 1 shows a summary of how the system works.

A. *Data Collection and Preprocessing Automation*

The primary data in this study were RSS feeds from Indonesian-language news websites, specifically Detik, CNBC Indonesia, CNN Indonesia, Tempo, and Kumparan. The data were collected through automated web scraping and immediately processed through a series of preprocessing steps involving the date and title fields. The preprocessing stages applied in this research include converting the date into the RFC 882 format, case folding, tokenizing, stemming, and removing stop words from the title and description text.

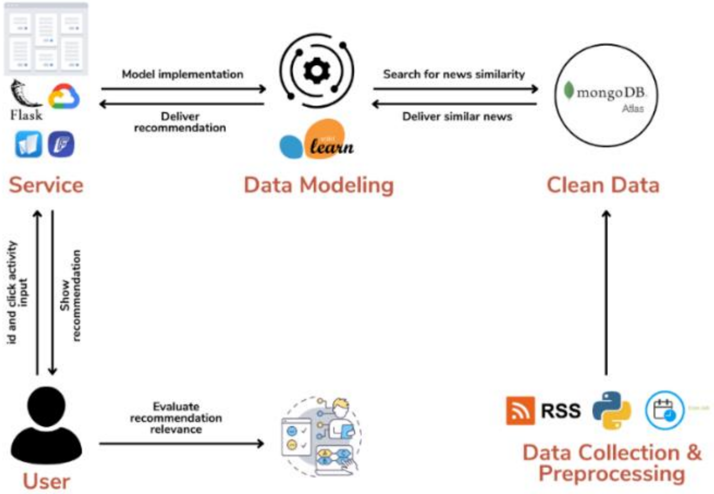


Figure 1 System Overview

1) *Convert Date to RFC 882*

This format is required as the standard date input for RSS feeds. The conversion is performed to standardize all data source time zones from Western Indonesian Time (WIB) to Greenwich Mean Time (GMT). The date format follows the string pattern: "date, %d %b %Y %H:%M:%S".

2) *Case Folding*

Case folding is the process of converting all characters in the text to lowercase. Additionally, this step removes non-alphabetic characters such as numbers, punctuation marks, and symbols [8].

3) *Tokenizing*

Tokenizing refers to segmenting text into smaller units, commonly words or phrases, known as tokens. This segmentation typically involves splitting alphabetic or alphanumeric sequences based on non-alphanumeric characters, such as punctuation or spaces [8].

4) *Stemming*

Stemming is the process of reducing words to their root form. In the Indonesian language context, stemming involves removing affixes, including suffixes, prefixes, and circumfixes [9].

5) *Stop Words Removal*

Stop words removal is the process of eliminating words that carry little to no meaningful value in a description by comparing them against a predefined stop list. The remaining words after this process are considered relevant keywords [9].

An example of the data scraping result and preprocessing output is presented respectively in Table 1 and Table 2.

TABLE I
SCRAPING RESULT EXAMPLE

Column	Contents
Title	Bertemu 2 Jam, Jaksa Agung-Kepala PPATK Banyak Terima Arahan dari Prabowo
Description	Jaksa Agung bersama jajaran, Kepala PPATK, hingga Plt Kepala BPKP bertemu dengan Presiden Prabowo di Istana Merdeka, Jakarta. Prabowo banyak memberikan arahan.
Link	https://news.detik.com/berita/d-7730747/bertemu-2-jam-jaksa-agung-kepala-ppatk-banyak-terima-arahan-dari-prabowo
Date	13 Jan 2025 17:38:49

The data scraping process obtains data fields such as Title, Description, Link, and Date to be used in the automated preprocessing, data modelling, and RSS feed generation process. RSS feeds typically provide a short description of the news article.

TABLE II
PREPROCESSING RESULT EXAMPLE

Result Column	Before Preprocessing	After Preprocessing
Title	Bertemu 2 Jam, Jaksa Agung-Kepala PPATK Banyak Terima Arahan dari Prabowo	temu 2 jam jaksa agung-kepala ppatk banyak terima arah prabowo
Date	13 Jan 2025 17:38:49	Mon, 13 Jan 2025 10:38:49 -0000

The data scraping and preprocessing results are stored in a MongoDB database. The entire process is also automated using a cron job to ensure periodic execution. The amount of data that has been obtained is around 54,821 (obtained from October 2024 to June 2025).

B. *Recommendation System*

A recommendation system is an information filtering mechanism designed to generate personalized item suggestions for users within an environment capable of collecting and storing data. One of the fundamental approaches to generating a recommendation list is to calculate content similarity or group content based on historical data—an approach commonly known as content-based filtering.

The algorithm employed in content-based filtering aims to measure the similarity of items previously rated or interacted with by the user. One of the most widely used algorithms in this context is TF-IDF, which quantifies the importance of terms in a corpus, often used in conjunction with cosine similarity to measure the closeness between items based on their feature vectors [10].

1) *TF-IDF*

TF-IDF is a text-weighting method that uses frequency-based data to assign weights to each term within a document. This weighting process is applied to text that has undergone preprocessing steps such as case folding, tokenization, stemming, and stop word removal. By applying TF-IDF, the feature representation of the text becomes more informative and accurate [11]. The calculation of TF-IDF weights is performed using the Python scikit-learn library. The mathematical formulation of TF-IDF according to the scikit-learn implementation is expressed as follows.

$$TF_{(t,d)} = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \qquad (1)$$

- $f_{t,d}$ is the frequency of term *t* in document *d*
- $\sum_{t' \in d} f_{t',d}^{'}$ is the total number of terms *t* across all documents *d*

$$IDF_{(t)} = \log \frac{1+n}{1+df(t)} + 1 \qquad (2)$$

- n is the total number of documents in the corpus
- $df(t)$ is the number of documents in which term *t* appears

2) Cosine Similarity

Cosine similarity is a method used to calculate the similarity between two text documents. The concept involves computing the cosine of the angle between two document vectors. The formula for cosine similarity, according to the scikit-learn implementation, is expressed as follows

$$Similarity_{(q,dj)} = \frac{q \cdot d_j}{|q||dj|} \quad (3)$$

- $|q|$ is the vector magnitude of the query
- $|d_j|$ is the vector magnitude of document *j*

$Similarity_{(q,dj)}$ represents the cosine similarity score between the query vector and the document vector. In this context, the query (representing previously visited news articles) is modeled as vector q, while the document (representing the collection of available news articles) is modeled as vector dj. Terms are extracted from the text corpus or database.

The computation of TF-IDF and cosine similarity is iterative, based on the number of news articles clicked by the user. The resulting recommendations from the computation process are stored in a MongoDB database along with corresponding user data. This entire process is automated using a cron job, allowing periodic execution.

3) User Profiling Approach

User profiling in this study is based solely on click behaviour, representing a form of implicit feedback. Implicit feedback can be used to infer relative preferences [12]. This method has been applied in [4], where the user's news category interest was predicted based on click behaviour. However, the current system does not capture additional behavioral indicators such as dwell time, which may provide more accurate signals of user engagement. The absence of such data limits the depth of user profiling and personalization. As a potential enhancement, future versions of the system could integrate explicit preference input, such as asking for news topics or categories preferences to support more accurate and adaptable recommendation generation.

C. Service Development

The RSS feed customization service was designed using the Flask library. Flask is a micro-framework for developing web-based application systems written in Python. It is widely recognized for its simplicity, flexibility, and ease of use [13]. In this study, Flask was utilized for the backend components of the RSS feed customization service.

Implementing Flask to support the customized RSS feed service, integrated with a recommendation system, involves using multiple routes (endpoints) along with their respective request types. The architectural design of the RSS feed customization service is illustrated in Figure 2.

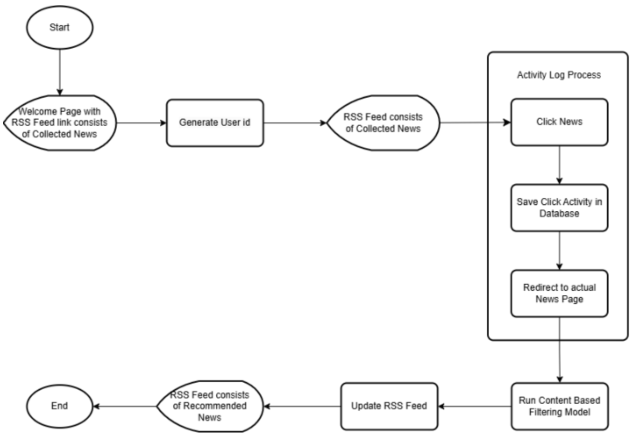

Start
Welcome Page with RSS Feed link consists of Collected News
Generate User id
RSS Feed consists of Collected News
Activity Log Process
Click News
Save Click Activity in Database
Redirect to actual News Page
Run Content Based Filtering Model
Update RSS Feed
RSS Feed consists of Recommended News
End

Figure 2. Service Development

D. Evaluation Approach of Service and Recommendation System

Based on the system design, which incorporates a content-based filtering recommendation model into the RSS feed customization process, evaluation and testing are necessary to assess the system's performance and user acceptance. User testing and satisfaction surveys were conducted to evaluate the relevance of the recommendations and the usability of the service from the user's perspective.

Due to the absence of ground truth data (such as predefined relevant articles or historical user click data), the use of standard evaluation metrics like precision, recall, or MAP was not applicable in this study. In such cases, a user study can be used as an alternative, particularly in early-stage recommendation system development. A user study is conducted by recruiting a (small) set of human test subjects who perform several pre-defined tasks that require them to interact with the recommendation system to provide quantitative and qualitative feedback. User studies particularly provide comprehensive feedback compared to the other experiment types and measure user experience at the time of recommendation [12].

This approach allows evaluators to judge the perceived relevance of news recommendations based on subjective experience, which is valuable when system usage data is limited. Similar user-centered evaluations have been used in related work [14] and [15], where they implemented human evaluation for a web-based recommendation system through a survey-based evaluation that was Customer Satisfaction Score (CSAT). This survey was conducted to evaluate the level of user satisfaction with the recommendation results provided by the system [15].

Therefore, the survey-based evaluation in this study focuses on aspects such as perceived recommendation relevance, usability, and user satisfaction, providing insight into how well the system meets user expectations in the absence of quantitative performance benchmarks.

1) *User Acceptance Testing (UAT)*

UAT is conducted as part of the evaluation phase. UAT is a process in which users test the system to verify that it meets their needs and expectations before being officially deployed. It serves as a documented confirmation that the developed software has been accepted by its users [16].

This study evaluates UAT across four key aspects: functionality, recommendation relevance, usability, and efficiency and productivity. Responses were collected using a Likert scale (ranging from 1 Strongly Disagree to 5 Strongly Agree) and analysed using the UAT percentage formula to assess the extent to which the service is accepted by users.

2) *CSAT*

CSAT is employed to measure user satisfaction with the service. CSAT is a simple yet effective method for quantifying satisfaction levels. As discussed by [17] and [14], CSAT measurement involves collecting user responses, calculating the number of positive responses, and applying those results to the CSAT formula to determine the satisfaction score.

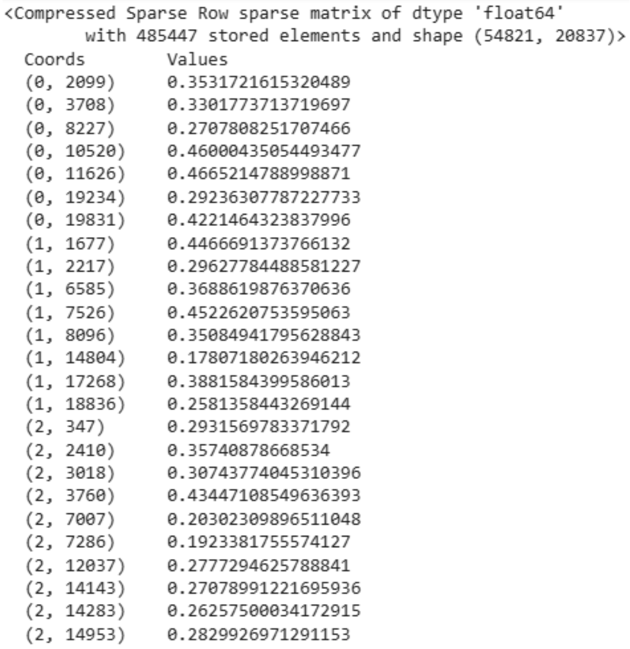
III. Result and Discussion

The result of the RSS feed customization service using a content-based filtering recommendation system is presented, starting from the system implementation phase and going through to the results and discussion.

A. TF-IDF and Cosine Similarity Calculation Result

Data modeling was implemented with the assistance of the sci-kit-learn library. The content-based filtering implementation involved computing TF-IDF weights and cosine similarity between user-clicked news articles and the entire data collection. An overview of the TF-IDF weights calculation result for the entire preprocessed news title data is shown in Figure 3.

The resulting TF-IDF matrix has dimensions of 54,821 × 20,837, indicating that it contains 54,821 document rows and 20,837 feature name columns. The command `print(tfidf_vectorizer)` was used to output the TF-IDF values without rendering the full matrix (which would contain a large number of zeros). The first column in the TF-IDF output refers to the index of the document, represented by a news title. The second column corresponds to the index of the feature name, which represents each word appearing across the document set.



```
<Compressed Sparse Row sparse matrix of dtype 'float64'
        with 485447 stored elements and shape (54821, 20837)>
  Coords        Values
  (0, 2099)     0.3531721615320489
  (0, 3708)     0.3301773713719697
  (0, 8227)     0.2707808251707466
  (0, 10520)    0.46000435054493477
  (0, 11626)    0.4665214788998871
  (0, 19234)    0.29236307787227733
  (0, 19831)    0.4221464323837996
  (1, 1677)     0.4466691373766132
  (1, 2217)     0.29627784488581227
  (1, 6585)     0.3688619876370636
  (1, 7526)     0.4522620753595063
  (1, 8096)     0.35084941795628843
  (1, 14804)    0.17807180263946212
  (1, 17268)    0.3881584399586013
  (1, 18836)    0.2581358443269144
  (2, 347)      0.2931569783371792
  (2, 2410)     0.35740878668534
  (2, 3018)     0.30743774045310396
  (2, 3760)     0.43447108549636393
  (2, 7007)     0.20302309896511048
  (2, 7286)     0.1923381755574127
  (2, 12037)    0.2777294625788841
  (2, 14143)    0.27078991221695936
  (2, 14283)    0.26257500034172915
  (2, 14953)    0.2829926971291153
```

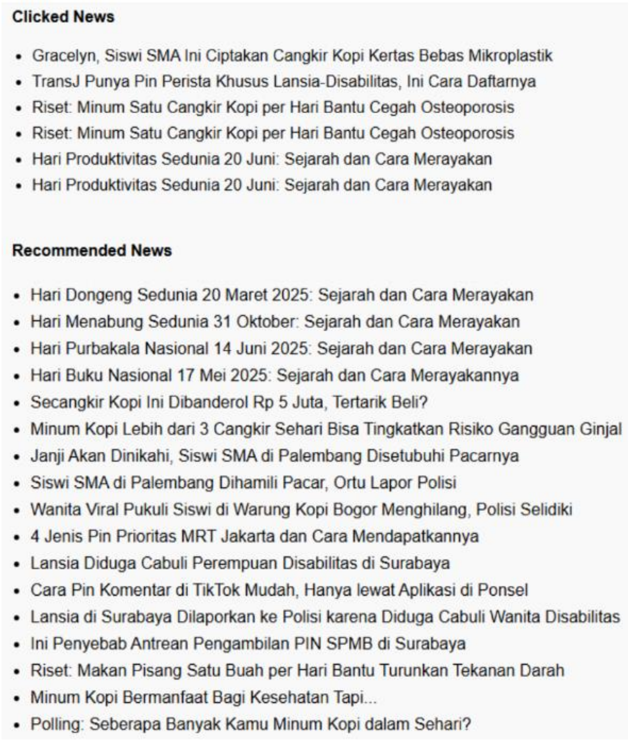
Figure 3. TF-IDF Weights for All Preprocessed News Title Data

These coordinates indicate the TF-IDF weight values, which typically range between approximately 0.1 and 0.4. Higher TF-IDF scores indicate the terms that are both frequent within the entire news title data and relatively rare across the entire corpus, suggesting they are good indicators of the news title's content. An example of a cosine similarity calculation result is shown in Table 3.

TABLE III
Scraping Result Example

Input	Recommendations	Similarity
hari produktivitas dunia 20 juni sejarah cara raya	Hari Pengungsi Sedunia 20 Juni 2025: Ini Tema dan Cara Merayakannya	0.595288
	Hari Dongeng Sedunia 20 Maret 2025: Sejarah dan Cara Merayakan	0.544775
	Hari Menabung Sedunia 31 Oktober: Sejarah dan Cara Merayakan	0.476902
	Hari Purbakala Nasional 14 Juni 2025: Sejarah dan Cara Merayakan	0.47267
	Hari Buku Nasional 17 Mei 2025: Sejarah dan Cara Merayakannya	0.392642

The news recommendation results are based on the clicked news item. The system recommends the top 5 news titles with the highest similarity to the clicked item. The cosine similarity values for the recommended articles range approximately from 0.3 to 0.5. An example of a news recommendation result for more than 1 clicked news item is shown in Figure 4.



Clicked News

- Gracelyn, Siswi SMA Ini Ciptakan Cangkir Kopi Kertas Bebas Mikroplastik
- TransJ Punya Pin Perista Khusus Lansia-Disabilitas, Ini Cara Daftarnya
- Riset: Minum Satu Cangkir Kopi per Hari Bantu Cegah Osteoporosis
- Riset: Minum Satu Cangkir Kopi per Hari Bantu Cegah Osteoporosis
- Hari Produktivitas Sedunia 20 Juni: Sejarah dan Cara Merayakan
- Hari Produktivitas Sedunia 20 Juni: Sejarah dan Cara Merayakan

Recommended News

- Hari Dongeng Sedunia 20 Maret 2025: Sejarah dan Cara Merayakan
- Hari Menabung Sedunia 31 Oktober: Sejarah dan Cara Merayakan
- Hari Purbakala Nasional 14 Juni 2025: Sejarah dan Cara Merayakan
- Hari Buku Nasional 17 Mei 2025: Sejarah dan Cara Merayakannya
- Secangkir Kopi Ini Dibanderol Rp 5 Juta, Tertarik Beli?
- Minum Kopi Lebih dari 3 Cangkir Sehari Bisa Tingkatkan Risiko Gangguan Ginjal
- Janji Akan Dinikahi, Siswi SMA di Palembang Disetubuhi Pacarnya
- Siswi SMA di Palembang Dihamili Pacar, Ortu Lapor Polisi
- Wanita Viral Pukuli Siswi di Warung Kopi Bogor Menghilang, Polisi Selidiki
- 4 Jenis Pin Prioritas MRT Jakarta dan Cara Mendapatkannya
- Lansia Diduga Cabuli Perempuan Disabilitas di Surabaya
- Cara Pin Komentar di TikTok Mudah, Hanya lewat Aplikasi di Ponsel
- Lansia di Surabaya Dilaporkan ke Polisi karena Diduga Cabuli Wanita Disabilitas
- Ini Penyebab Antrean Pengambilan PIN SPMB di Surabaya
- Riset: Makan Pisang Satu Buah per Hari Bantu Turunkan Tekanan Darah
- Minum Kopi Bermanfaat Bagi Kesehatan Tapi...
- Polling: Seberapa Banyak Kamu Minum Kopi dalam Sehari?

Figure 4. Recommendation Result Example in Database

The TF-IDF and cosine similarity calculations were performed iteratively based on the number of clicked news articles (n), producing a total of a maximum of n × 5 recommended articles. This approach was adopted to ensure that the recommendations could reflect the diversity of topics from the clicked articles. However, it is necessary to reassess the number of recommended articles, as the total may become excessive if many articles have been clicked. The cosine similarity score distribution based on the news recommendation results from users' click activity in the UAT and CSAT survey is presented in Figure 5.

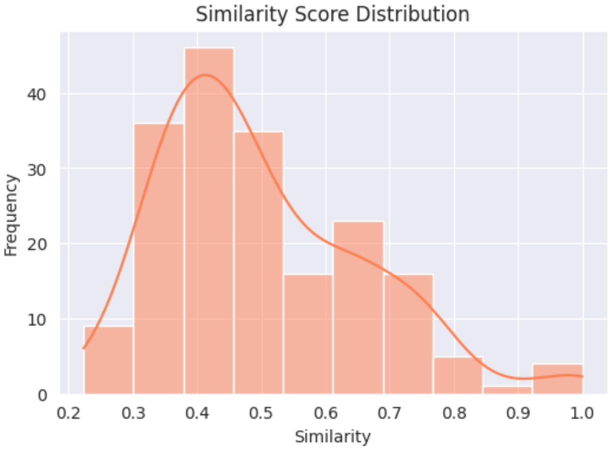


Figure 5. Cosine Similarity Score Distribution

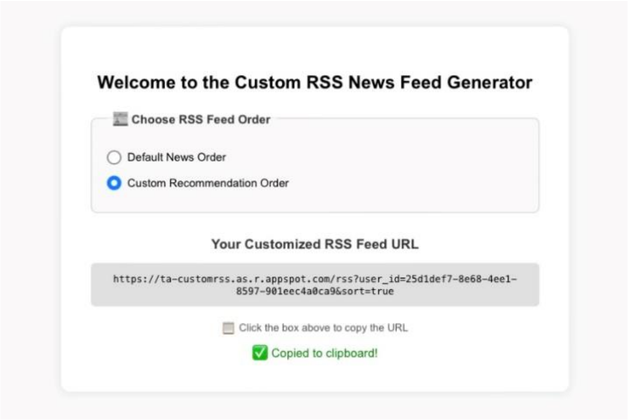
Most similarity values fall within the range of approximately 0.3 to 0.5. These values indicate that the similarity between the recommended news titles and the clicked news titles is generally low to moderate. This outcome is likely due to the varying values within the TF-IDF matrix across all recommended news titles, which results in lower similarity with the TF-IDF matrix of the clicked news. Interestingly, a few recommended articles yield a cosine similarity score of 1.0, indicating that the titles are identical to the clicked news item.

Evaluation metrics commonly used in recommendation systems, such as precision, recall, or MAP, were not employed in this study due to the lack of ground truth data within the dataset. As a result, offline evaluation of recommendation relevance could not be performed. Therefore, relevance assessment was carried out through a user study in the UAT survey, as an alternative method for assessing how well the recommended articles matched user interests. The relevance evaluation score distribution is later presented in the *UAT Result*. This approach is adopted because the service is still under development, and sufficient user interaction data is not yet available to support quantitative, metric-based evaluation.

Based on the conducted experiments, the majority of recommendation results were able to present relatively relevant news articles. This may be attributed to the presence of numerous articles with titles that closely resemble the topic of the clicked news. Moreover, the structure of the recommended news titles also tends to be similar to the structure of the clicked news titles, which aligns well with the study's objective of generating content with high surface-level similarity. However, since the weighting method used (TF-IDF) emphasizes term frequency and importance, the similarity between news articles is limited to those aspects. The semantic meaning of the content cannot be captured using TF-IDF alone. Therefore, alternative weighting methods such as LSI, Doc2Vec, BERT, or deep learning approaches like CNNs or news encoders may be explored in future work to generate more personalized recommendations based on user interaction patterns or underlying topics and named entities within the news content.

B. Service Development Result

The data collection and modeling results were implemented in a Flask-based application to design a service for RSS feed customization based on news recommendation outcomes. The application includes several endpoints, each designed to perform specific processes. These processes include accessing the welcome page, tracking news-click activities, generating and accessing RSS feeds, and updating recommendations and RSS feed content. The implementation was effective, as it could execute all these processes and produce accurate results. Figure 6 shows the service's welcome page that was implemented with HTML and CSS.



Welcome to the Custom RSS News Feed Generator

Choose RSS Feed Order

○ Default News Order

● Custom Recommendation Order

Your Customized RSS Feed URL

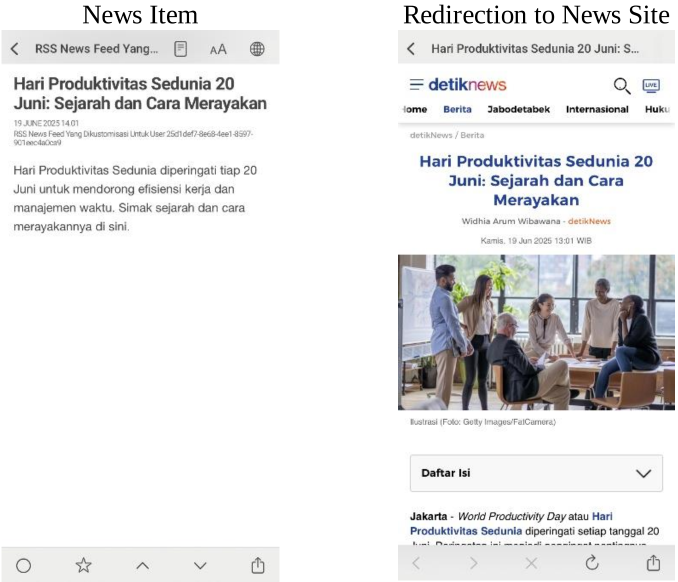
```
https://ta-customrss.as.r.appspot.com/rss?user_id=25d1def7-8e68-4ee1-8597-901eec4a0ca9&sort=true
```

□ Click the box above to copy the URL

✓ Copied to clipboard!

Figure 6. RSS Feed Customization Service's Welcome Page

The welcome page serves to display the RSS feed URL to new users. Each time this page is accessed, a unique user ID is generated. The service provides 2 features for the RSS feed order type. The Default News Order will not show the news recommendation result, and the Custom Recommendation Order will generate a new set of timestamps to mark the news recommendations' time. To access the RSS feed, users can copy the URL provided in the placeholder and paste it into an RSS feed reader by clicking the designated link. To obtain news recommendations, the user has to read at least 1 news item by clicking the news item and ensuring they are redirected to the actual news site. An example of this process is illustrated in Figure 7.



News Item

RSS News Feed Yang...

Hari Produktivitas Sedunia 20 Juni: Sejarah dan Cara Merayakan

19 JUNE 2025 14.01

RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8e68-4ee1-8597-901eec4a0ca9

Hari Produktivitas Sedunia diperingati tiap 20 Juni untuk mendorong efisiensi kerja dan manajemen waktu. Simak sejarah dan cara merayakannya di sini.

Redirection to News Site

Hari Produktivitas Sedunia 20 Juni: S...

detikNews / Berita

Hari Produktivitas Sedunia 20 Juni: Sejarah dan Cara Merayakan

Widhia Arum Wibawana - detikNews

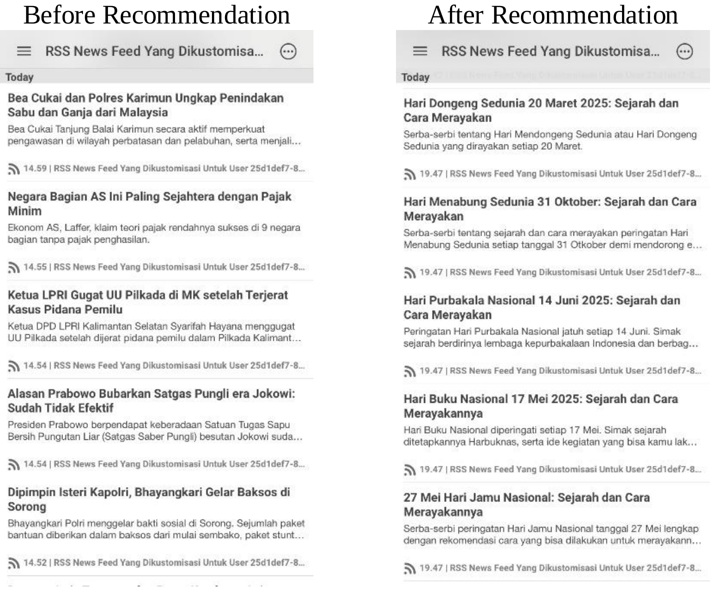
Kamis, 19 Jun 2025 13:01 WIB

Ilustrasi (Foto: Getty Images/FatCamera)

Daftar Isi

Jakarta - *World Productivity Day* atau **Hari Produktivitas Sedunia** diperingati setiap tanggal 20

Figure 7. Activity Log Process



Before Recommendation

RSS News Feed Yang Dikustomisa...

Today

Bea Cukai dan Polres Karimun Ungkap Penindakan Sabu dan Ganja dari Malaysia
Bea Cukai Tanjung Balai Karimun secara aktif memperkuat pengawasan di wilayah perbatasan dan pelabuhan, serta menjali...
14.59 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Negara Bagian AS Ini Paling Sejahtera dengan Pajak Minim
Ekonom AS, Laffer, klaim teori pajak rendahnya sukses di 9 negara bagian tanpa pajak penghasilan.
14.55 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Ketua LPRI Gugat UU Pilkada di MK setelah Terjerat Kasus Pidana Pemilu
Ketua DPD LPRI Kalimantan Selatan Syarifah Hayana menggugat UU Pilkada setelah dijerat pidana pemilu dalam Pilkada Kalimant...
14.54 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Alasan Prabowo Bubarkan Satgas Pungli era Jokowi: Sudah Tidak Efektif
Presiden Prabowo berpendapat keberadaan Satuan Tugas Sapu Bersih Pungutan Liar (Satgas Saber Pungli) besutan Jokowi suda...
14.54 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Dipimpin Isteri Kapolri, Bhayangkari Gelar Baksos di Sorong
Bhayangkari Polri menggelar bakti sosial di Sorong. Sejumlah paket bantuan diberikan dalam baksos dari mulai sembako, paket stunt...
14.52 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

After Recommendation

RSS News Feed Yang Dikustomisa...

Today

Hari Dongeng Sedunia 20 Maret 2025: Sejarah dan Cara Merayakan
Serba-serbi tentang Hari Mendongeng Sedunia atau Hari Dongeng Sedunia yang dirayakan setiap 20 Maret.
19.47 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Hari Menabung Sedunia 31 Oktober: Sejarah dan Cara Merayakan
Serba-serbi tentang sejarah dan cara merayakan peringatan Hari Menabung Sedunia setiap tanggal 31 Otkober demi mendorong e...
19.47 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Hari Purbakala Nasional 14 Juni 2025: Sejarah dan Cara Merayakan
Peringatan Hari Purbakala Nasional jatuh setiap 14 Juni. Simak sejarah berdirinya lembaga kepurbakalaan Indonesia dan berbag...
19.47 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

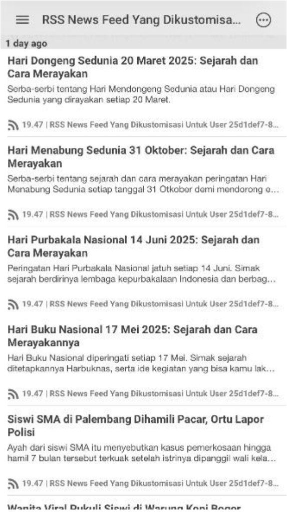
Hari Buku Nasional 17 Mei 2025: Sejarah dan Cara Merayakannya
Hari Buku Nasional diperingati setiap 17 Mei. Simak sejarah ditetapkannya Harbuknas, serta ide kegiatan yang bisa kamu lak...
19.47 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

27 Mei Hari Jamu Nasional: Sejarah dan Cara Merayakannya
Serba-serbi peringatan Hari Jamu Nasional tanggal 27 Mei lengkap dengan rekomendasi cara yang bisa dilakukan untuk merayakann...
19.47 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Figure 8. Before and After RSS Feed Customization Comparison

When a user clicks on a news item, they are redirected to the original news page, and the information about the clicked article is stored in the database. Subsequently, the user receives news recommendations through their RSS feed reader application in the next update cycle, which occurs approximately two hours after the initial click. Figure 8 presents a comparison of the recommendation results displayed in the RSS feed, as viewed using the Newsify RSS feed reader mobile application.

The RSS feed display after a user clicks on a news item (illustrated in Figure 8) shows a list of recommended articles related to national and international observances, which appear starting from the sixth position in the feed to still provide default news data for the first 5 news item. This recommendation display is only visible during the next RSS feed update cycle. The newly recommended items are marked with a timestamp labeled “Today”, indicating that the recommendations were generated on the same day as the click activity. Although the recommended articles remain stored in the RSS feed, their timestamps do not update in subsequent feed refreshes. As a result, the articles are later marked with older timestamps, such as “1 day ago” (as shown in Figure 9), which causes them to shift to lower positions in the RSS feed.


RSS News Feed Yang Dikustomisa...

1 day ago

Hari Dongeng Sedunia 20 Maret 2025: Sejarah dan Cara Merayakan
Serba-serbi tentang Hari Mendongeng Sedunia atau Hari Dongeng Sedunia yang dirayakan setiap 20 Maret.
19.47 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Hari Menabung Sedunia 31 Oktober: Sejarah dan Cara Merayakan
Serba-serbi tentang sejarah dan cara merayakan peringatan Hari Menabung Sedunia setiap tanggal 31 Otkober demi mendorong e...
19.47 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Hari Purbakala Nasional 14 Juni 2025: Sejarah dan Cara Merayakan
Peringatan Hari Purbakala Nasional jatuh setiap 14 Juni. Simak sejarah berdirinya lembaga kepurbakalaan Indonesia dan berbag...
19.47 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Hari Buku Nasional 17 Mei 2025: Sejarah dan Cara Merayakannya
Hari Buku Nasional diperingati setiap 17 Mei. Simak sejarah ditetapkannya Harbuknas, serta ide kegiatan yang bisa kamu lak...
19.47 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Siswi SMA di Palembang Dihamili Pacar, Ortu Lapor Polisi
Ayah dari siswi SMA itu menyebutkan kasus pemerkosaan hingga hamil 7 bulan tersebut terkuak setelah istrinya dipanggil wali kela...
19.47 | RSS News Feed Yang Dikustomisasi Untuk User 25d1def7-8...

Wanita Viral Pukuli Siswi di Warung Kopi Bogor

Figure 9. News Recommendation Contents After 1 Day

Based on the service development results, an issue was identified regarding inconsistency in the RSS feed customization behavior. Recommended articles appear correctly at the top section of the feed (starting from position six) immediately after the next update cycle. However, during subsequent updates, the recommendation timestamps are no longer refreshed, causing the items to be pushed further down the feed despite the recommendation time update function being triggered. This issue is likely due to the caching mechanism of the RSS feed reader, which may prevent re-reading or updating the publication time for identical articles, particularly when the article is generated based on previous click activity.

C. Service and Recommendation System Evaluation

The system was evaluated based on the implementation of the RSS feed customization service developed using news recommendation results. Two user study methods were employed: UAT and the CSAT survey, which was tested by 20 respondents and conducted online using their devices.

1) UAT Result

In this study, UAT was conducted by distributing a step-by-step usage guide for the service to users, followed by a questionnaire to be completed after they used the service.

The results of the UAT evaluation showed that all assessed aspects achieved an average percentage score above 80%. The recommendation relevance aspect scored 92%as the second-highest scoring aspect. Figures 10 and 11 illustrate the score distribution for recommendation relevance based on clicked news and recommendation compatibility with the respondent's preference, respectively.

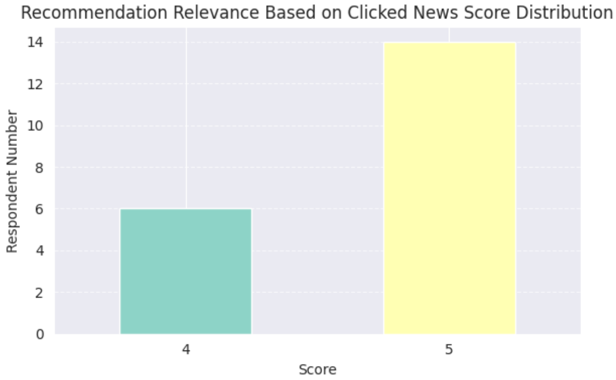


Figure 10. Score Distribution for Recommendation Relevance Based on Clicked News

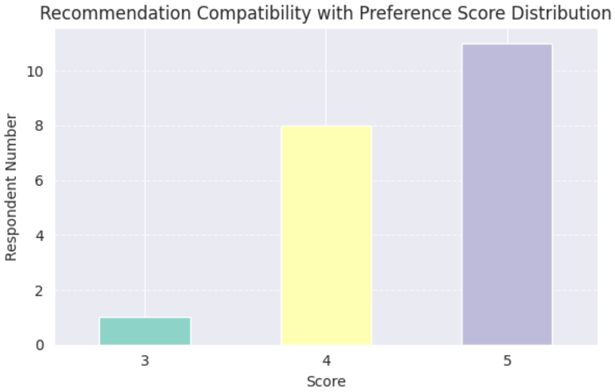


Figure 11. Score Distribution for Recommendation Alignment with Respondent’s Personal Interest

The majority of respondents (14 individuals) strongly agreed that the recommended news articles were relevant to the articles they had clicked. A portion of the respondents (11 individuals) strongly agreed that the recommended articles aligned with their interests. Additionally, 8 respondents agreed, while 1 respondent remained neutral regarding the relevance of the recommendations to their interests. This neutral response was given despite the user selecting a political news article, which typically aligns with their usual news consumption. This may be attributed to the entities or topics within the recommended articles not fully reflecting the respondent's specific interests. These score distributions show that almost all respondents agree that the news recommendations are relevant to the clicked news and align with their interests.

Based on the UAT score interpretation in [16], the previously mentioned percentage results place all aspects within the "Excellent" category of the UAT assessment scale, indicating that users were able to effectively accept and utilize the service, particularly concerning the recommendation outcomes.

2) CSAT Result

CSAT was used to measure user satisfaction with the developed service. Like the UAT process, CSAT was conducted by distributing usage instructions followed by a user questionnaire.

The results of the CSAT regarding customization result and overall experience each received a 90% or higher CSAT score, which qualifies as "Excellent CSAT" according to the American Customer Satisfaction Index (ASCI) in [14]. Furthermore, the overall satisfaction score was 92.5%, which was also categorized as "Excellent CSAT".

These evaluation results indicate that the designed service—especially the recommendation output—is well-received by users, with the majority expressing high satisfaction with their experience using the service.

IV. CONCLUSION

The design of a customized RSS feed service for collecting and recommending news based on content similarity consists of several stages: data collection via web scraping, data preprocessing, content-based filtering modeling, and RSS feed design and customization. Data scraping and preprocessing were automated using a cron job. The content-based filtering model was implemented using TF-IDF and cosine similarity with the support of the scikit-learn library. The RSS feed design and customization were developed in a Flask-based web application organized through multiple endpoints. The recommendations generated based on user click interactions were reasonably relevant, as they successfully presented news titles similar to the clicked articles, with cosine similarity scores ranging from 0.2 to 1.0. However, several challenges were encountered, including inconsistencies in the caching mechanisms within the RSS feed reader application.

The customized RSS feed service was tested through a user study. The testing yielded highly satisfactory results: all aspects of the UAT achieved an average percentage score above 80%, indicating an "Excellent" rating, while the overall CSAT reached 90% or above, which falls within the "Excellent CSAT" range. These results are valid, as most respondents were technologically literate and consumed news near-daily.

The absence of ground truth data presents a limitation in applying traditional evaluation metrics such as precision, recall, or nDCG. Future work could incorporate explicit relevance judgments or implicit behavioral data (e.g., dwell time, long-term clicks) to enable quantitative evaluation metrics.

Future studies should also consider addressing caching mechanisms within RSS feed reader applications. Although the recommendation timestamp was successfully updated, the reader application appears to retain cached data, which prevents it from consistently displaying the most recent content. It is recommended to conduct compatibility testing with other RSS reader applications or consider technical approaches such as modifying the `<guid>` tag to improve update detection.

REFERENCES

[1] KBBI, "Kamus Besar Bahasa Indonesia (KBBI) Daring." Accessed: Dec. 25, 2024. [Online]. Available: https://kbbi.kemdikbud.go.id/

[2] Builtwith, "RSS Usage Statistics." Accessed: Jan. 07, 2025. [Online]. Available: https://trends.builtwith.com/feeds/RSS

[3] Y. Elyusufi and E. M. En-Naimi, "User profiling and RSS feeds recommendation based on the ontological approach," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Nov. 2017. doi: 10.1145/3167486.3167496.

[4] J. Liu, P. Dolan, and E. R. Pedersen, "Personalized News Recommendation Based on Click Behavior," *IUI'10*, 2010.

[5] M. Pranji´c, V. Podpečan, M. Robnik-ˇ Sikonja, S. Pollak, and J. Stefan, "Evaluation of related news recommendations using document similarity methods," in *Conference on Language Technologies & Digital Humanities*, 2020. [Online]. Available: https://github.com/vpodpecan/lemmagen3

[6] P. Sitikhu, K. Pahi, P. Thapa, and S. Shakya, "A Comparison of Semantic Similarity Methods for Maximum Human Interpretability," in *2019 Artificial Intelligence for Transforming Business and Society (AITB)*, Kathmandu, Nepal, 2019, pp. 1–4. doi: 10.1109/AITB48515.2019.8947433.

[7] G. Yunanda, D. Nurjanah, and S. Meliana, "Recommendation System from Microsoft News Data using TF-IDF and Cosine Similarity Methods," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, Jun. 2022, doi: 10.47065/bits.v4i1.1670.

[8] F. Ramadhan and A. Musdholifah, "Online Learning Video Recommendation System Based on Course and Sylabus Using Content-Based Filtering," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 3, p. 265, Jul. 2021, doi: 10.22146/ijccs.65623.

[9] A. Sanjaya, A. Bagus Setiawan, U. Mahdiyah, I. Nur Farida, A. Risky Prasetyo, and U. Nusantara PGRI Kediri, "Pengukuran Kemiripan Makna Menggunakan Cosine Similarity dan Basis Data Sinonim Kata Measurement of Meaning Similarity Using Cosine Similarity and Word Synonyms Database," vol. 10, no. 4, 2023, doi: 10.25126/jtiik.2023106864.

[10] A. H. J. P. Juni Permana and Agung Toto Wibowo, "Movie Recommendation System Based on Synopsis Using Content-Based Filtering with TF-IDF and Cosine Similarity," *International Journal on Information and Communication Technology (IJoICT)*, vol. 9, no. 2, pp. 1–14, Dec. 2023, doi: 10.21108/ijoict.v9i2.747.

[11] I. Riadi, A. Fadlil, and U. Ahmad Dahlan Yogyakarta, “Identifying Hate Speech in Tweets with Sentiment Analysis on Indonesian Twitter Utilizing Support Vector Machine Algorithm,” *Jurnal Ilmu Komputer dan Informatika*, vol. 9, no. 2, 2023.

[12] E. Zangerle and C. Bauer, “Evaluating Recommender Systems: Survey and Framework,” *ACM Comput. Surv.*, vol. 55, no. 8, 2022.

[13] A. S. Ningrum and A. W. Utami, “Sistem Deteksi Tulisan Tangan Aksara Hangeul Satu Silabel Menggunakan Metode Convolutional Neural Network (CNN) Berbasis Framework Flask,” *JEISBI*, vol. 04, no. 04, 2023.

[14] K. S. Yanisa Putri, I Made Agus Dwi Suarjaya, and Wayan Oger Vihikan, “Sistem Rekomendasi Skincare Menggunakan Metode Content Based Filtering dan Collaborative Filtering,” *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 4, no. 3, pp. 764–774, Aug. 2024, doi: 10.51454/decode.v4i3.601.

[15] Ni Putu Adnya Puspita Dewi, D. P. S. Putri, and I. N. P. Trisna, “Fine-Tuned Transformer Models for Keyword Extraction in Skincare Recommendation Systems,” *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 3, p. 966, 2025, [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[16] Aliyah Aliyah, Nahrun Hartono, and Asrul Azhari Muin, “Penggunaan User Acceptance Testing (UAT) Pada Pengujian Sistem Informasi Pengelolaan Keuangan Dan Inventaris Barang,” *Switch : Jurnal Sains dan Teknologi Informasi*, vol. 3, no. 1, pp. 84–100, Dec. 2024, doi: 10.62951/switch.v3i1.330.

[17] A. Astutik, S. A. A. Yusuf, A. Rahman, N. Sulistianingsih, and M. A. J. Akbar, “Penguatan Literasi Digital Dengan Family Link Bagi PKK Desa Sesela: Workshop Pengawasan Penggunaan Media Sosial Keluarga Yang Aman & Produktif,” *Jurnal Pengabdian kepada Masyarakat Nusantara (JPkMN)*, vol. 5, no. 3, 2024.