e-ISSN : 2600-8793

Available online at
https://jcrinn.com/
https://crinn.conferencehunter.com/

Journal of Computing Research and Innovation 10(1) 2025

Journal of Computing Research and Innovation

Encouraging Recycling in Bangi Selatan Through a Content-Based Filtering Web Application

Muhammad Akmal Ahmad Nawawi[1], Tajul Rosli Razak[1*]

[1,2] College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Malaysia

ARTICLE INFO

Article history:
Received 24 October 2024
Revised 31 December 2024
Accepted 10 January 2025
Online first
Published 1 March 2025

Keywords:
Recommendation system
Content-based filtering
Information retrieval
Recycling awareness

DOI:
10.24191/jcrinn.v10i1

ABSTRACT

This study addresses the challenges faced by residents of Bangi Selatan in adopting 3R (Reduce, Reuse, Recycle) practices, primarily due to a lack of interest in conservation efforts and insufficient awareness of recycling's importance. To address these challenges, we presented a web application that enhances recycling adoption by delivering personalized content recommendations. The key contributions of this study include the development of a novel recommendation system based on content-based filtering (CBF) with improved accuracy through a modified Term Frequency-Inverse Document Frequency (TF-IDF) formula. We compare various recommendation techniques, including collaborative and hybrid filtering, and demonstrate how CBF effectively improves user engagement with recycling content. Our methodology involves advanced text vectorization and cosine similarity for precise content matching. User acceptance testing confirms the system's effectiveness in increasing user engagement with relevant recycling information. This study highlights the potential of personalized recommendation systems in promoting environmental conservation and provides a foundation for future enhancements in recycling initiatives.

1. INTRODUCTION

Artificial Intelligence (AI) has transformed various sectors with models like ChatGPT and BardAI, showcasing large language models' power to enhance global efficiency (Gill et al., 2024). However, despite these technological strides, the issue of proper solid waste management remains pressing, especially in developing countries. Global waste production is expected to rise from 1.3 billion tonnes to over 2.2 billion by 2025 (David et al., 2019). Despite 80% of this waste being recyclable, much is in landfills due to poor waste management practices (Zaharudin et al., 2023).

[2*] Corresponding author. *E-mail address*: tajulrosli@uitm.edu.my
https://doi.org/10.24191/jcrinn.v10i1.510

In Malaysia, the recycling rate stands at 33.7%, proving that recycling practices remain low (Tiew et al., 2019). Based on the survey conducted in the Bangi Selatan area, 52.9% of respondents do not recycle because they were not encouraged due to a lack of exposure. The residents believed that traditional recycling campaigns were insufficient in promoting recycling awareness because they lacked engagement with recycling education. To address these gaps, leveraging a content-based filtering (CBF) system can be pivotal in encouraging recycling by personalizing recommendations based on user preferences and interests.. Such systems make them ideal for promoting recycling by delivering targeted information on recyclable materials, which can significantly enhance user engagement and foster sustainable recycling habits.

This study proposes the development of a CBFweb application designed to enhance recycling exposure in Bangi Selatan. The system utilizes personalized recommendations powered by cosine similarity and the TF-IDF (Term Frequency-Inverse Document Frequency) algorithm. To tailor content to user interests, a vital feature of the application is a search box that enables users to retrieve relevant information from the database. Raising awareness and improving local recycling rates could be a model for integrating technology and environmental sustainability through user-centered solutions.

2. LITERATURE

This section overviews the factors behind recycling awareness shortfalls, the opinions of Bangi Selatan locals, the development of web applications and the recommendation system.

2.1 Factors Behind Recycling Awareness Shortfalls

The researcher identifies several factors of recycling awareness shortfalls like insufficient information exposure, inadequate 3R (reduce, reuse, recycle) practices, ineffective campaigns by the government, etc (Kate, 2022; Tiew et al., 2019). This assertion is supported by Gilani et al. (2023) findings, where their research identified a lack of awareness of environmental care as the leading cause of why plastic waste is the most waste disposal in landfills. When the importance of recycling is not taken seriously, it will lead to diminished public commitment to 3R practices over time.

2.2 Opinion Of Bangi Selatan Locals

A survey of Bangi Selatan locals revealed that a need for more environmental awareness is the main issue for neglecting 3R practices. Furthermore, respondents firmly believe that education programs play a significant role in understanding recycling benefits and addressing the shortfalls. While the impact of the low resale value of recyclables is acknowledged, it is considered a less significant factor, with only 41.2% of respondents citing it as a reason for reduced recycling activity. Besides that, respondents state that the distance between their homes and the recycling centres causes their efforts not to recycle.

2.3 Type of Web Application

Web applications are server-based programs accessible via browsers such as Google Chrome and Microsoft Edge (Contributor, 2023). They are built using HTML, CSS, and JavaScript, forming the framework for their development together. These applications operate through client-server communication: when a user (the client) requests data via an HTTP message over TCP/IP, the server responds by sending the requested data in packets (MDN, 2023). Among various types of web applications, there are eight groups of web applications, including static websites, dynamic web applications, e-commerce, single-page web applications, animated web applications, content management systems (CMS) web applications and rich internet applications. Choosing the correct type of web application is crucial to ensure it is relevant to the project. Table 1 below compares three types of web applications that have been chosen.

https://doi.org/10.24191/jcrinn.v10i1.510

Table 1. Comparison of the types of web applications

Characteristics	Static Web Application	Dynamic Web Application	CMS Web Application
Technology	Client-side only	Client-side and server-side	Depends on plugins and template
Interactivity	Depends	Depends	Responsive
Functionality	Limited	Real-time content	Depends on plugins and template
Coding Skills	Moderate	High	Low
Development Cost	Cheap	Expensive	Expensive license
Pros	Speed Performance and reliability	Can add database and many features	SEO-friendly and time-efficient
Cons	Non-dynamic content	Higher development time	Limited customization
Technology	Client-side only	Client-side and server-side	Depends on plugins and template

Recommendation systems (RS) are tools that retrieve information based on user interactions and interests from a large pool of data, suggesting content aligned with their activities and preferences (Liling, 2019; Maphosa & Maphosa, 2023). RS is used widely around the globe, especially in entertainment streaming like Netflix and Amazon Prime Video. According to Maphosa & Maphosa (2023), RS are innovative tools that predict what users might do next by looking at things they like and have done before from historical data.

2.4 Overview Of Recommendation System

Content-based Filtering (CBF) is a fundamental RS that suggests the content by analysing relationships between user interests and item representation (Liling, 2019). CBF depends on the similarity of the user profile and items to suggest the user's potential content (Liling, 2019; Maphosa & Maphosa, 2023; Pujahari & Sisodia, 2022). Both past behaviours from the user profile and item representation are used to build a user preference profile; later, the user preference will be used as a dataset to compare new items with the user's historical behaviour (Pujahari & Sisodia, 2022). One advantage of CBF is its capability to recommend relevant information to a user even when there has not been any prior interaction between the user and the system (Pérez-Almaguer et al., 2021).

Next, a collaborative filtering (CF)-based recommendation system relies on gathering user behavior data to determine item relevance based on user preferences or similarities, ultimately generating recommendations by finding associations between users with comparable interests (Liling, 2019; Maphosa & Maphosa, 2023; Pujahari & Sisodia, 2022). In CF, one widely employed method is the rating system, which gauges user satisfaction or preference regarding a recommended item (Fkih,2023). Unlike CBF, CF requires interactions among users to establish connections between similar user profiles, making it highly effective in environments with rich user interaction data.

Additionally, hybrid recommendation systems aim to overcome the limitations of both CBF and CF by integrating multiple recommendation technologies. This approach leverages the strengths of each method, enhancing recommendation accuracy and compensating for data limitations. By doing so, hybrid systems effectively address challenges like the cold start problem commonly encountered in CF systems (Liu et al., 2022). For instance, in a movie recommendation application, the Content-Based Collaborative Filtering (CBCF) model first utilizes CBF to analyse item attributes and generate initial user preferences. These preferences are then refined using CF to identify similarities among users with analogous tastes, resulting in more personalized recommendations (Afoudi et al., 2021).

https://doi.org/10.24191/jcrinn.v10i1.510

3. METHODOLOGY

This section discusses the specific approach employed to accomplish the research objectives by highlighting several important methods implemented in the CBF algorithm construction.

3.1 Data Classification

For item representation, the valuable information related to recycling content data is stored in the string, and every string represents the description of the content material. The content material comprises 20 examples, with five focused on e-waste, three on paper, six on plastics, three on metals, and three on general recycling. This collection includes pollution facts, awareness information, and fun facts about recycling, all designed to highlight the environmental importance of recycling. Since the user profile is based on an input query, therefore data sample for the user profile is insignificant.

3.2 Text Vectorization

In order to determine the similarity between the user query and content item representation, it is essential first to convert the text into numerical vectors. The Term Frequency-Inverse Document Frequency (TF-IDF) technique plays a crucial role in this transformation, making it possible to compare user queries and content descriptions meaningfully. This approach is widely adopted in content-based filtering (CBF) systems. Given that the proposed system is centred around text representation and information retrieval, the content and user queries are represented as numerical vectors, where each term in the content description is assigned, a value reflecting its relevance. Thus, effective for the CBF system relies on user query input as the user profile representation. TF-IDF calculates the significance of each term by considering two key factors:

(i) Term frequency (TF)

To determine the importance of a term within a content description or user query, TF calculates how often a term appears, with more frequent terms indicating higher relevance. This approach highlights the critical terms for more accurate content recommendations. TF formula is provided as follows:

$$TF = \frac{n_{i,j}}{\sum_k n_{i,j}} \quad (1)$$

Where n refers to the number of occurrences of terms in content, i denotes the specific term, j denotes specific content, and k denotes all terms in the content.

(ii) Inverse document frequency (IDF)

The standard IDF formula has been modified to incorporate a smoothing technique in this system. This adjustment serves two essential purposes: first, it prevents excessively high IDF values that could arise with the standard formula; second, it ensures that terms that do not appear in any document are still assigned a nonzero and non-negative IDF value. By balancing term frequency with document frequency, this modified IDF improves the system's ability to make accurate content recommendations to its users. The modification theorem is as follows:

$$IDF = 1 + \log\left(\frac{N+1}{df_i + 1}\right) \quad (2)$$

where N represents the total content and df_i is the number of terms mentioned in the content.

https://doi.org/10.24191/jcrinn.v10i1.510

(iii) TF-IDF

The TF-IDF process results in two key vectors: one for the content item and one for the user query. These vectors are derived using the TF-IDF formula, TF-IDF = TF × IDF, which provides a numerical representation of term relevance in both the content and user input.

3.3 Similarity Computation

Cosine similarity measures the similarity between the TF-IDF vector representation of the item representation (content description) and the TF-IDF vector representation of user interest (user query). Theoretically, a smaller cosine distance indicates that the similarity between the two nonzero vectors is closer and more relevant to recommend. The cosine similarity theorem is shown below:

$$\cos\theta = \frac{A \cdot B}{||A|| \cdot ||B||} \quad (3)$$

A and B are nonzero vectors, representing user interests and item representation respectively.

3.4 Content-based Filtering Architecture and Algorithm

The CBF architecture for this system is illustrated in Figure 1, which presents a model tailored to handle search queries within a user interface. The core functionality of this model is to recommend content that aligns with the user's search input by leveraging a string-based similarity computation. When a user enters a query into the search box, this input is treated as the user's profile. The model then compares this profile against the textual representations of available content, which consists of materials related to recycling awareness and education. By analysing the similarity between the user's query and the content descriptions, the system can suggest relevant materials that closely match the user's interests, thereby providing a personalized and efficient information retrieval experience.

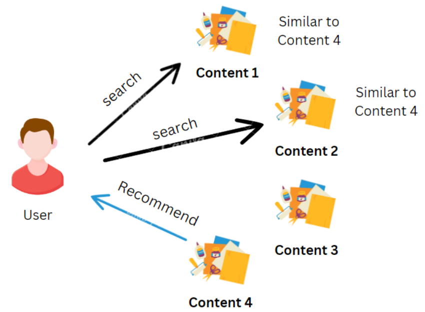


Fig 1: Proposed Content-based Filtering Model

In realizing this model, the system employs a feature extraction using the TF-IDF method to transform the query and the content into numerical vectors. These vectors are essential for measuring the degree of similarity between the user's input and the available content using cosine similarity. This technique allows the system to quantify the relevance of each content item to the user's query, ensuring that the most pertinent materials are recommended.

4. RESULT AND DISCUSSIONS

This section shows the experimental evaluation of the modification of the IDF theorem. The test will conduct to determine whether the proposed CBF algorithm manages to recommend to users the recycling

https://doi.org/10.24191/jcrinn.v10i1.510

content based on their input query, including its recommendation accuracy. Besides that, acceptance testing was conducted to assess whether the proposed system could effectively promote interest in 3R practices among residents of Bangi Selatan.

4.1 Evaluation Result

The performance of the proposed IDF theorem and standard theorem are summarized in Table 2, highlighting three document terms, which are "awareness', "recycling", and "plastics" from 20 documents (content) as the dataset.

Table 2. Comparison of IDF value between different theorems

Document Terms	awareness	plastics	recycling
Standard Theorem	1.2040	1.3863	0.2231
Proposed Theorem	2.2528	2.0986	1.2113

Based on Table 2 above, the Standard Theorem yields IDF values of 1.2040, 0.2231, and 1.3869 for these terms, while the Proposed Theorem produces higher values of 2.2528, 1.2113, and 2.0986. Consequently, the Proposed Theorem consistently generates higher IDF values, especially for standard terms, potentially assigning greater weight to frequent terms in subsequent calculations. This adjustment could lead to improved outcomes compared to the Standard Theorem. This highlights how choosing the Smoothing technique can substantially influence the relevance and weighting of terms, potentially enhancing the overall performance of the recommendation system. Moreover, Tables 3 and 4 below present the results, demonstrating the system's ability to recommend recycling content, specifically when incorporating the user query input.

Table 3. Example of user input and content material

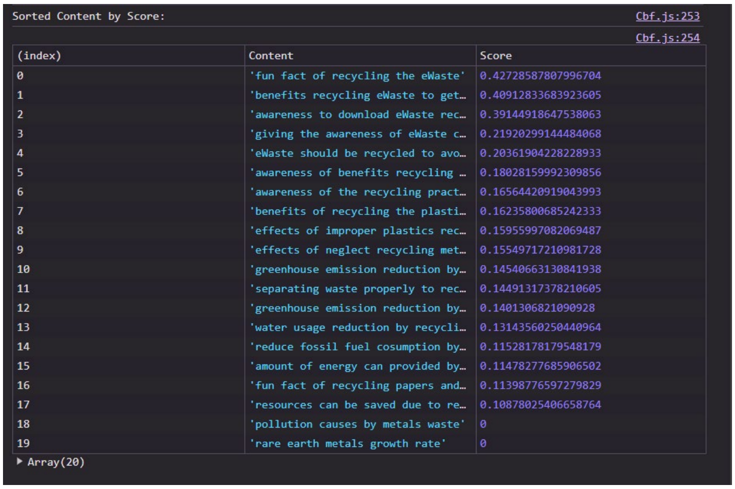
ID	Vector	Document Text/User Input
1	Item representation	Greenhouse emission reduction by recycling papers
2	Item representation	e-Waste should be recycled to avoid pollution
3	User Interest	papers dump effect

Table 4. Similarity and content recommendation with ID=3

Vector	1	2
Similarity	Yes	No
Recommend to user	Yes	No
Reason	Have a paper term	No matched term
Cosine similarity value	Exist with user query	Does not exist with user query

The experiment results in Table 4 illustrate the effectiveness of the proposed CBF system in recommending relevant content to users based on their input. Using cosine similarity, we compare the user input (ID=3) with two document vectors (ID=1 and ID=2). Document ID=1 contains the term "papers," which also exists in the user input, leading to a positive cosine similarity value. As a result, this document is recommended to the user. In contrast, document ID=2 lacks any matching terms with the user input, resulting in a negative similarity assessment and the document not being recommended. This demonstrates that the system successfully identifies relevant documents when standard terms are present between the input and content, ensuring accurate recommendations.

https://doi.org/10.24191/jcrinn.v10i1.510


Sorted Content by Score: Cbf.js:253

Cbf.js:254

(index)	Content	Score
0	'fun fact of recycling the eWaste'	0.42728587807996704
1	'benefits recycling eWaste to get…	0.40912833683923605
2	'awareness to download eWaste rec…	0.39144918647538063
3	'giving the awareness of eWaste c…	0.21920299144484068
4	'eWaste should be recycled to avo…	0.20361904228228933
5	'awareness of benefits recycling …	0.18028159992309856
6	'awareness of the recycling pract…	0.16564420919043993
7	'benefits of recycling the plasti…	0.16235800685242333
8	'effects of improper plastics rec…	0.15955997082069487
9	'effects of neglect recycling met…	0.15549717210981728
10	'greenhouse emission reduction by…	0.14540663130841938
11	'separating waste properly to rec…	0.14491317378210605
12	'greenhouse emission reduction by…	0.1401306821090928
13	'water usage reduction by recycli…	0.13143560250440964
14	'reduce fossil fuel cosumption by…	0.11528178179548179
15	'amount of energy can provided by…	0.11478277685906502
16	'fun fact of recycling papers and…	0.11398776597279829
17	'resources can be saved due to re…	0.10878025406658764
18	'pollution causes by metals waste'	0
19	'rare earth metals growth rate'	0

▶ Array(20)

Fig. 2: Proposed Content-based Filtering Model

Next, Figure 2 presents the results of the CBF recommendation algorithm when the user query includes the topics "e-waste" and "recycling." The system successfully ranks content items based on their cosine similarity scores, with the most relevant content appearing at the top. The scores range from this maximum value down to 0, clearly indicating content relevance. Items with a score of 0 were deemed irrelevant to the user's query. This distribution of similarity scores indicates that the system can effectively distinguish between highly relevant, moderately relevant, and irrelevant content. As a result, users are presented with a well-curated selection of information tailored to their specific interests, encouraging them to practice recycling through reading.

5. USER ACCEPTANCE TESTING

The user acceptance survey conducted in Bangi Selatan, involving 30 respondents, showed positive feedback regarding the CBF web system's impact on recycling behaviour. Users reported feeling more engaged in recycling after reading the recommended content tailored to their interests based on their input queries. The CBF system, powered by TF-IDF and cosine similarity, successfully matched most user inputs to relevant content. Additionally, respondents indicated that the web application was impactful in encouraging recycling, showing overall positive user acceptance. However, issues were raised regarding the system's inability to recognise synonymous terms, such as "recycling" and "recycled," which occasionally hindered its effectiveness in delivering relevant recommendations. Besides that, some users expressed frustration with the system's handling of unknown inputs, noting that when it could not recognise their interests, it provided alternative content that was often unrelated. Addressing this limitation could enhance the system's usability and accuracy in future iterations.

6. CONCLUSION AND RECOMMENDATION

In conclusion, the outcome of User Acceptance Testing (UAT) with residents in Bangi Selatan successfully met the project's objectives. Based on the current evidence, the online survey results indicate that the content-based filtering recommendation system effectively promotes recycling practices by delivering tailored recycling-related content through the search box. This approach has significantly increased user awareness and understanding of recycling's importance.

https://doi.org/10.24191/jcrinn.v10i1.510

However, the implementation of TF-IDF with cosine similarity presents some limitations. TF-IDF's focus on word frequency rather than semantic meaning can lead to missed connections between synonyms, such as "paper" and "papers." Additionally, the system may face issues with sparsity as vocabulary size grows, diminishing the effectiveness of vectorization and leading to decreased cosine similarity scores. To enhance the system's accuracy and usability, future iterations should consider incorporating semantic analysis to address synonym recognition and reduce sparsity challenges.

7. ACKNOWLEDGEMENTS

I want to extend my most profound appreciation to my supervisor, Dr Tajul Rosli bin Razak, for their invaluable guidance, unwavering support, and insightful feedback throughout this journal writing. I also thank God for granting me the opportunity to pursue my bachelor's in computer science and for the strength to complete this challenging endeavor. I am thankful to the academic staff of Universiti Teknologi MARA (UiTM) for their assistance and resources and to my family and friends for their patience, understanding, and constant motivation during difficult times. Special thanks to the Rakan Alam Sekitar UiTM Student Chapter for their invitation, which enriched my experience in environmental conservation. Finally, I am deeply grateful to all the participants who shared their time and experiences, making this study possible.

8. REFERENCES

Afoudi, Y., Lazaar, M., & Al Achhab, M. (2021). Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network. Simulation Modelling Practice and Theory, 113. https://doi.org/10.1016/j.simpat.2021.102375

Contributor, T. (2023b, January 20). web application (web app). Software Quality. https://www.techtarget.com/searchsoftwarequality/definition/Web-application-Web-app

David, A., Thangavel, Y. D., & Sankriti, R. (2019). Recover, recycle and reuse: An efficient way to reduce the waste. International Journal of Mechanical and Production Engineering Research and Development, 9(3), 31–42. https://doi.org/10.24247/ijmperdjun20194

Fkih, F. (2023). Enhancing item-based collaborative filtering by users' similarities injection and low-quality data handling. Data & Knowledge Engineering, 144, 102126. https://doi.org/10.1016/j.datak.2022.102126

Gilani, I. E., Sayadi, S., Zouari, N., & Al-Ghouti, M. A. (2023). Plastic waste impact and biotechnology: Exploring polymer degradation, microbial role, and sustainable development implications. In Bioresource Technology Reports (Vol. 24). Elsevier Ltd. https://doi.org/10.1016/j.biteb.2023.101606

Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., … Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. Internet of Things and Cyber-Physical Systems, 4, 19–23. https://doi.org/10.1016/j.iotcps.2023.06.002

How the web works - Learn web development | MDN. (2023, November 17). MDN Web Docs. https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/How_the_Web_works

Liling, L. (2019). Summary of recommendation system development. Journal of Physics: Conference Series, 1187(5). https://doi.org/10.1088/1742-6596/1187/5/052044

Liu, J., Shi, C., Yang, C., Lu, Z., & Yu, P. S. (2022). A survey on heterogeneous information network based recommender systems: Concepts, methods, applications and resources. AI Open, 3, 40–57. https://doi.org/10.1016/j.aiopen.2022.03.002

Maphosa, V., & Maphosa, M. (2023). Fifteen Years of Recommender Systems Research in Higher Education: Current Trends and Future Direction. In Applied Artificial Intelligence (Vol. 37, Issue 1). Taylor and Francis Ltd. https://doi.org/10.1080/08839514.2023.2175106

Pérez-Almaguer, Y., Yera, R., Alzahrani, A. A., & Martínez, L. (2021). Content-based group recommender systems: A general taxonomy and further improvements. Expert Systems with Applications, 184, 115444. https://doi.org/10.1016/j.eswa.2021.115444

Pujahari, A., & Sisodia, D. S. (2022). Item feature refinement using matrix factorization and boosted learning based user profile generation for content-based recommender systems. Expert Systems with Applications, 206. https://doi.org/10.1016/j.eswa.2022.117849

Tiew, K. G., Basri, N. E. A., Deng, H., Watanabe, K., Zain, S. M., & Wang, S. (2019). Comparative study on recycling behaviours between regular recyclers and non regular recyclers in Malaysia. In Journal of Environmental Management (Vol. 237, pp. 255–263). Academic Press. https://doi.org/10.1016/j.jenvman.2019.02.033

Zaharudin, Z. A., Shuib, A., Hadianti, R., & Rodzi, Z. M. (2023). Towards Sustainable City: A Covering Model for Recycling Facility Location-allocation in Nilai, Malaysia. Science and Technology Indonesia, 8(4), 570–578. https://doi.org/10.26554/sti.2023.8.4.570-578