Jurnal Sains Informatika Terapan (JSIT)
E-ISSN: 2828-1659, Volume: 4, Issue: 3, Year: 2025
Page: 700-705, Available online at: https://rcf-indonesia.org/home/

IMPLEMENTATION OF TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) ALGORITHM ON HALAL NETWORK INTERNATIONAL (HNI) PRODUCT SEARCH FEATURE

Annud Auliani Hasanah[1)], Ninik Agustin[2)]), M.Khanif[3]
[1,2,3] Universitas Nahdlatul Ulama Al Ghazali
Cilacap
Corresponding Author: [1] annudaul@gmail.com, [2] ninik.agustin@unugha.id

Article Info

Article history:

Received: Juny,12, 2025
Revised: Sept 15, 2025
Accepted: okt 30, 2025
Published: des 30, 2025

Keywords:

HNI Products
Search Features
TF-IDF Algorithm
React.js

ABSTRACT

Halal Network International (HNI) products already have around 100 products which are divided into three categories, namely Herbs Products, Health Food & Beverages, and Cosmetics & Home Care. These products are published through an electronic catalog that provides product information and search features to facilitate users. The mismatch between keywords and product descriptions causes the search to fail to display relevant results. To overcome this problem, the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm is implemented in the search feature. This algorithm is used to calculate the weight of each word in the product description and determine the relevance of the keywords entered by the user. The system is built with React.js framework on the frontend, Node.js on the backend, and PostgreSQL as the database. Based on the test results, the average precision value is 71%, recall 85% and accuracy 73%. These results show that the TF-IDF algorithm is effective in improving the relevance of product search results.

1. INTRODUCTION

HNI (Halal Network International) is a company that focuses its business on halal and quality products. HNI products are halal-certified by MUI and marketed through the Sharia MLM system[1]. There are currently around 100 products divided into 3 categories, namely Herbs products, Health Food & Beverages, and Cosmetics & Home Care. Herbs Product includes herbal medicines that are made from plants and do not use chemicals[2]. HNI's product information is published in an electronic catalog so that it can be accessed by anyone with internet access[3]. In the electronic catalog, there is a search feature that is used to facilitate product search.

However, the search feature in the electronic catalog still has limitations, especially for ordinary users who are not familiar with product names and benefits. Errors in writing product names cause the search to fail, and the system is not able to recommend relevant products. Therefore, it is necessary to improve the search feature by implementing the TF-IDF (Term Frequency-Inverse Document Frequency) Algorithm. This algorithm is able to give weight to each word in the document, so that the search can be more accurate and relevant.

Several studies have explored search feature improvements. The combination of the TF-IDF algorithm with the query expansion method improved search efficiency in an e-catalog system [4]. A thesis title search system was developed using the Django framework and PostgreSQL, with testing using the System Usability Scale (SUS) yielding a score of 77.17, indicating good usability[5]. A comparison between TF-IDF and SQL Query algorithms in an archive search system showed that while SQL Query was faster in processing time, TF-IDF provided better relevance of search results[6]. The TF-IDF algorithm was also applied to search Indonesian National Standard (SNI) documents, enabling users to find documents using general keywords[7]. In a drug search system, TF-IDF and cosine similarity were used to match drugs based on disease indications, and showed good performance after preprocessing and weighting steps[8]. The combination of SVM, TF-IDF, and Word2Vec was able to classify library books with good accuracy (F1-Score 76%)[9]. TF-IDF and Haversine were applied to a photographer recommendation system, with 78% of respondents

stating that the system worked well[10]. TF-IDF and Cosine Similarity were used to classify thesis topics, resulting in effective clustering based on keyword similarity[11].

Based on this background, this research aims to apply the TF-IDF algorithm to the HNI product search feature and measure the performance of the search feature by testing accuracy, precision, recall and f1 score.

2. MATERIALS AND METHODS

The research procedure is a series of systematic steps in compiling the research from the beginning to the end of the research, as described in Figure 1.

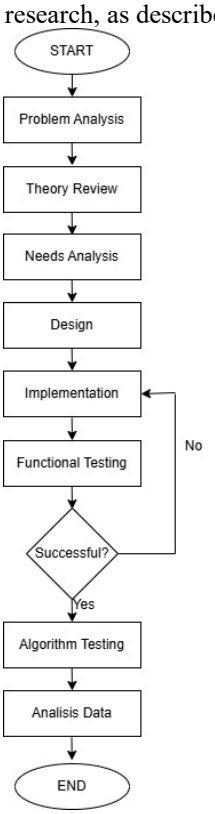

Figure 1 Research Flowchart Diagram

2.1. Problem Analysis

Root cause analysis is a technique to find out the root of all the problems that occur [12]. This analysis is carried out by determining the main problems which are then made causal relationships that aim to find effective solutions..

2.2. Theory Review

After analyzing the problem, the researcher conducts a theoretical study which is used to provide a conceptual foundation and an in-depth view of the topic under study[13].

2.3. Needs Analysis

Requirements analysis is the process of collecting the necessary data and studying the data in depth[14]. The needs analysis includes system requirements and research needs shown in table 1. System needs analysis aims to analyze what needs are needed to build a system[12].

Table 1 User Requirements

No	User	Access Rights
1	Admin	Login
		Adding Data
		Deleting Data
		Changing Data
2	User	Entering Keyword
		View Search Results

Table 2 Software Requirements

No	Tools/Materials	Description
1	Windows 11 64-bit	Operating System used
2	Visual Studio Code	Text Editor
3	Google Chrome	Browser used to run the system
4	Draw.io	To create flowcharts and other diagrams
5	JavaScript	Front End programming language
6	Node JS	Back End programming language
7	PostgreSQL	Database
8	React JS	Framework

2.4. Design

The design stage aims to minimize errors when making a system that is designed using use case diagrams, sequence diagrams, class diagrams, activity diagrams, wireframes, and TF-IDF algorithm flow[15].

2.4.1. Use Case Digram

Use Case Diagram is one of the diagrams in UML that serves to describe the interaction relationship between actors (both users and external systems) and the system being analyzed[16]. System analysis will be made as in Figure 2 which consists of 2 actors, namely admin and user.

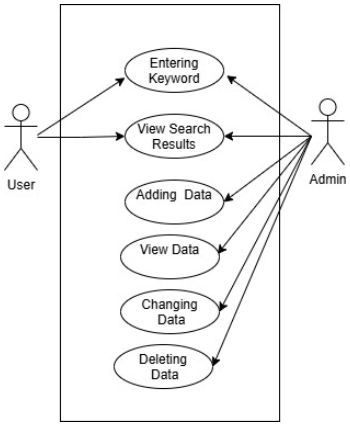

Figure 2 Use Case Diagram

2.4.2. Sequence Diagram

Sequence diagrams are used to explain interactions between objects in a system in detail[17].

2.4.3. Class Diagram

Class diagrams are used to clearly illustrate the structure of a system, including classes, attributes, methods and relationships between classes[18].

2.4.4. Activity Diagram

Activity diagrams are used to describe the flow of information in the system. the following is an activity diagram of the system to be created[19].

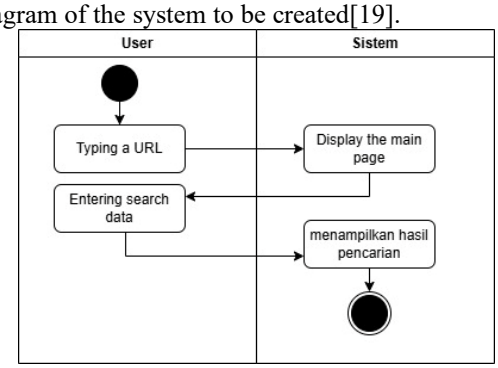

Figure 3 Activity Diagram

2.4.5. Wireframe

A wireframe is a simple overview of the system's layout and structure, showing the placement of key elements on a page[20]. The wireframe of this website is as follows Figure 4 illustrates the main page of the website, users can use the search feature.

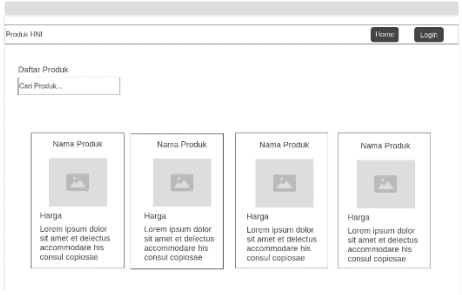

Figure 4 Wireframe

2.4.5. TF-IDF Algorithm Flow

TF-IDF algorithm in this system as shown in Figure 5.

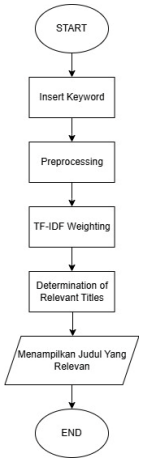

Figure 5 TF-IDF Algorithm Flow

3.5. Implementation

The implementation stage starts with text preprocessing which includes lowercasing, punctuation removal, tokenization, and empty element removal[21].

Furthermore, the calculation of TF (Term Frequency) and IDF (Inverse Document Frequency) is carried out to generate TF-IDF weights for each document[22]. The calculation of the TF-IDF algorithm is shown as follows :

$$tf_{t,d} = \frac{number\ of\ word\ occurrences}{\text{number of words in document}}$$

$$idf_t = \log\left(\frac{D}{df_t}\right)$$

$$W_{d,t} = tf_{t,d}\ x\ idf_t$$

Keterangan :

$W_{d,t}$ = Weight value of document d against word t

$tf_{t,d}$ = frequency of occurrence of word t in document d

D = total number of documents

df_t = number of documents containing word t

idf_t = inverse document frequency value of word t

3.6. Testing

Functional testing is carried out using black box testing, then algorithm testing is carried out by calculating accuracy, recall precision and f1 score. Precision measures the accuracy of the system in classifying positive or negative data, while recall shows the system's ability to retrieve relevant data. Accuracy represents the closeness of the system's predictions to human predictions[23]. F1 Score is a combination of precision and recall values to measure balance[24].

$$\text{Akurasi} : \frac{TP+TN}{TP+FP+FN+TN}$$

$$\text{Presisi} : \frac{TP}{TP+FP}$$

$$\text{Recall} : \frac{TP}{TP+FN}$$

$$\text{F1 Score} : \frac{2x\ (recall\ x\ presisi)}{(recall\ x\ presisi)}$$

Description:
TP (True Positive): Number of correct predictions for the positive class (products that are truly relevant to the search)
TN (True Negative): Number of correct predictions for the negative class (products that are not relevant to the search)
FN (False Negative): Number of incorrect predictions for the negative class (products that are relevant but not recognized by the model)
FP (False Positive) : Number of incorrect predictions for the positive class (products that were incorrectly considered relevant)

3. RESULTS AND DISCUSSION

At this stage the system is implemented using Frontend, Backend and Database. The following are the results of the implementation of the features in the system. The Home menu is the main page displayed when users access the website. In this system, the page is distinguished based on login status, namely general users and admins.

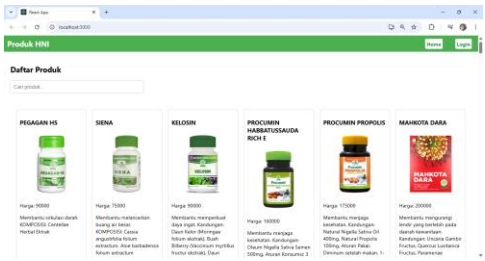

Figure 6 Home Menu Display

The Login menu is a feature that functions as an authentication gate for admins to be able to access product management features on the system.

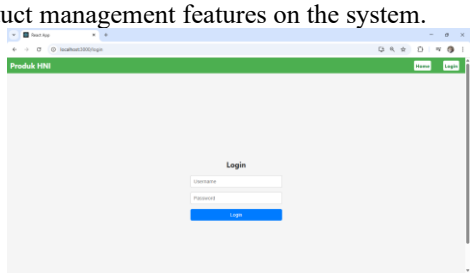

Figure 7 Login Menu Display

The Add Product menu is one of the features that can be accessed by the admin after the login process. This feature allows admins to add new product data to the system.

Figure 8 Add Product Menu Display

The product edit menu is very important to maintain the accuracy and relevance of product data in the catalog. Through this feature, admins can change product information, such as name, price, description, tags and upload product images.

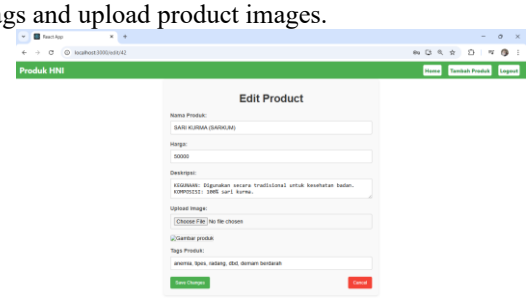

Figure 9 Product Edit Menu Display

The product delete feature serves to delete products that are no longer relevant or available from the system. On the admin home page, each product is equipped with a delete button. When the button is clicked, the system will automatically display a confirmation dialog as shown in the figure.

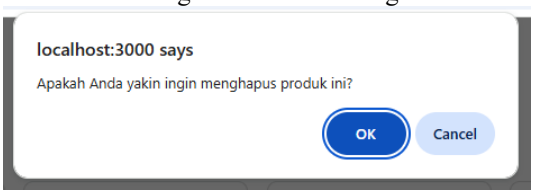

Figure 10 Product Delete Notification

3.1. Implementation of TF-IDF Algorithm

The TF-IDF (Term Frequency-Inverse Document Frequency) algorithm is implemented in the system to support the product search feature based on keywords. The search process starts when the user types the keywords in the search field on the main page. After the keywords are received by the server, preprocessing is performed on both the keywords and the entire document. Preprocessing is done with stages including, converting the entire text to lowercase, removing punctuation using regular expressions, breaking the text into words (tokenizing) and removing empty elements.

After all the text has gone through the preprocessing stage, the weighting process is carried out using the TF-IDF algorithm. The system will display by sorting the list of products based on the highest score value to the lowest.

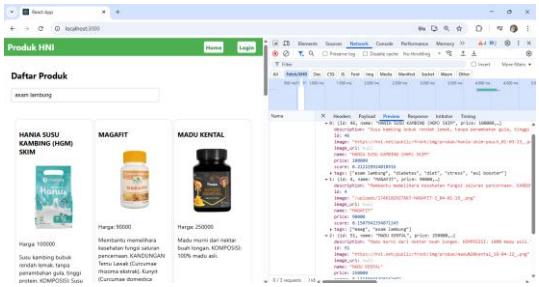

Figure 11 Product Search Result

Functional testing is carried out using Black Box Testing, which tests the functionality of the system based on input and output without paying attention to the internal structure of the program code[25].

3.2. Algorithm Testing

Tests were conducted using 34 keywords that represent user needs. After the user enters the keywords, the data predicted by the system is compared with the actual data. The following table presents the average evaluation results of all keywords used.

Tabel 1 Algorithm Testing Results

Evaluation Metrics	Average Value
Precision	71%
Recall	85%
Accuracy	73%
F1 Score	73%

From the test results on a number of varied keywords, the system showed an average precision value of 71% and an average recal of 85%. This shows that the system tends to be more effective in displaying many products relevant to the keywords entered by the user, although it still needs improvement in reducing the appearance of irrelevant results. In addition, an average accuracy value of 73% was obtained which shows that overall, the system has a fairly good ability to correctly identify search results. The average F1 Score value of 73% also supports this finding, where the metric describes the balance between precision and recall.

4. CONCLUSION

This research is motivated by the need for an efficient search system in the HNI product catalog, especially to help ordinary users find suitable products without having to know the exact product name. Term Frequency-Inverse Document Frequency (TF-IDF) algorithm is considered effective in calculating the weight of words in documents so as to increase the relevance of search results. This research is limited to the development of React JS-based search features that are only focused on HNI products..

The system is designed using the waterfall method and tested using Black Block Testing and algorithm performance evaluation through accuracy, precision and recall metrics. Based on the results of the research and discussion that has been carried out, the following conclusions can be drawn :

4.1. This research successfully implemented Term Frequency-Inverse Document Frequency (TF-IDF) algorithm on Halal Network International (HNI) product search feature. The system is able to display relevant products based on the keywords typed by the user, without the need to know the product name.

4.2. Based on tests using 34 keywords, an average precision value of 71%, an average recall of 85%, and an average accuracy of 73% were obtained. The results show that the system performs quite well in displaying products that are relevant to user needs.

REFERENCES

[1] Z. U. Nukha, R. R. Kurniati, and R. N. Hardati, "Pengaruh Sertifikasi Halal, Harga dan Kualitas Produk terhadap Kepuasan Konsumen Melalui Keputusan Pembelian," *Jiagabi*, vol. 10, no. 1, pp. 75–84, 2021, [Online]. Available: https://jim.unisma.ac.id/index.php/jiagabi/article/viewFile/9981/7912

[2] G. C. Putra, N. P. Pandawani, and M. E. A. Citra, "Peningkatan Kualitas Produk Herbal dan Kosmetika Natural Ba," *J. Bakti Sar.*, vol. 04, no. 02, pp. 91–100, 2015, [Online]. Available: https://media.neliti.com/media/publications/74815-ID-peningkatan-kualitas-produk-herbal-dan-k.pdf

[3] R. Wati, "Perancangan E-Katalog Berbasis Web Pada RR Collection Sampit Sebagai Media Branding Menggunakan Aplikasi Figma," *EJECTS E-Journal Comput. Technol. Informations Syst.*, vol. 2, no. 1, pp. 8–13, 2022.

[4] S. F. Kusuma, T. Badriyah, P. Wibowo, R. Faradisa, and S. Huda, "Optimalisasi Fitur Pencarian Pada E-Catalog Menggunakan Query Expansion Dan Algoritma TF-IDF," *Techno.Com*, vol. 22, no. 3, pp. 703–711, 2023, doi: 10.33633/tc.v22i3.8698.

[5] Ardianzah, Roni, and P. . Husni Thamrin, S.T, M.T, "Pengembangan Sistem Pencarian pada Aplikasi Skripsi untuk Meningkatkan Hasil Pencarian Judul," Universitas Muhammadiyah Surakarta, 2024. [Online]. Available: https://eprints.ums.ac.id/120772/

[6] D. Asmarajati, "Analisis Perbandingan Algoritma Tf-Idf Dengan Sql Query Untuk Kasus Pencarian Pada Sistem Informasi Dokumentasi Arsip (Sidokar)," *Device*, vol. 10, no. 1, pp. 1–8, 2020, doi: 10.32699/device.v10i1.1478.

[7] N. Ani, D. Y. Sinaga, N. Junior, and M. D. Munggaran, "Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF- IDF) Untuk Fitur Pencarian Dokumen Standar Nasional Indonesia," *J. Sci. Appl. informatics*, vol. 6, no. 3, pp. 517–522, 2023, [Online]. Available: https://jurnal.umb.ac.id/index.php/JSAI/article/view/6504

[8] K. Djunaidi, R. F. Ningrum, D. T. Kusuma, and K. M. Saradiba, "Pencarian Fungsional Obat Menggunakan Algoritma Tf-Idf Dan Cosine Similarity," *PETIRPengkajian dan Penerapan Tek. Inform.*, vol. 17, no. 1, pp. 77–83, 2024.

[9] S. N. Cahyani and G. W. Saraswati, "Implementation of Support Vector Machine Method in Classifying School Library Books With Combination of Tf-Idf and Word2Vec," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1555–1566, 2023, doi: 10.52436/1.jutif.2023.4.6.1536.

[10] A. Irvandani, "Sistem Rekomendasi Fotgrafer dengan Metode Haversine dan TF-IDF di Malang Raya," InstitutTeknologi Nasional Malang, 2020.

[11] H. Sari, G. L. Ginting, T. Zebua, and Mesran, "Penerapan Algoritma Text Mining dan TF-IDF Untuk Pengelompokan Topik Skripsi Pada Aplikasi Repository STMIK Budi Darma," *Terap. Inform. Nusant.*, vol. 2, no. 7, pp. 414–432, 2021, [Online]. Available: https://ejurnal.seminar-id.com/index.php/tin

[12] Y. Dewantara, "Perhitungan Kapasitas Baterai dan Arus

Komponen pada AR.Drone Quadcopter untuk Estimasi Waktu dan Jarak Terbang," Universitas Brawijaya Malang, 2011. [Online]. Available: https://repository.dinus.ac.id/docs/ajar/5a._handout

[13] D. F. D. N. Mulyono Apriyano, "Metodologi Penelitian Pertanian," *Metodol. Penelit. Pertan.*, pp. 1–10, 2021.

[14] W. R. Chaniago, R. Hafsari, R. A. Sari, and M. Ardiansyah, "Perancangan Sistem Informasi Inventory Berbasis Web pada PT. Bintang Oriental," *J. Sist. Inf.*, vol. 8, no. 5, pp. 1–6, 2024.

[15] M. Badrul, "Penerapan Metode waterfall untuk Perancangan Sistem Informasi Inventory Pada Toko Keramik Bintang Terang," *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 8, no. 2, pp. 47–52, 2021, doi: 10.30656/prosisko.v8i2.3852.

[16] P. R. Cahyani, L. Maylinasari, S. A. Ambami, and B. R. Putra, "Analisis Dan Desain Sistem Aplikasi Kantin Elektronik (E-Canteen) Bagi Mahasiswa Dan Staff Universitas," *J. Digit. Bus. Innov. Manag.*, vol. 2, no. 2, pp. 164–179, 2023, doi: 10.26740/jdbim.v2i2.58084.

[17] S. Sandfreni, M. B. Ulum, and A. H. Azizah, "Analisis Perancangan Sistem Informasi Pusat Studi Pada Fakultas Ilmu Komputer Universitas Esa Unggul," *Sebatik*, vol. 25, no. 2, pp. 345–356, 2021, doi: 10.46984/sebatik.v25i2.1587.

[18] M. S. I. Mustakim, Deyidi Mokoginta, Steiva Amerlien Sandra Wowiling and O. V. Indra, Ade Suparman, "Perancangan Sistem Informasi Penggajian Berbasis Web Dengan Metode Waterfall," *J. Cakrawala Inf.*, vol. 2, no. 1, pp. 59–68, 2022, doi: 10.54066/jci.v2i1.176.

[19] J. Simangunsong, A. Voutama, and H. Hannie, "Rancang Bangun Sistem Informasi Online Marketplace Berbasis Web Application," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1261–1268, 2023, doi: 10.36040/jati.v7i2.6793.

[20] E. U. P. T. Xyz, "PERANCANGAN WIREFRAME DAN MOCKUP SISTEM," vol. 7, no. 1, pp. 28–35, 2025, doi: 10.55642/eatij.v7i01.966.

[21] R. Al Rasyid and D. H. U. Ningsih, "Penerapan Algoritma TF-IDF dan Cosine Similarity untuk Query Pencarian Pada Dataset Destinasi Wisata," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 8, no. 1, pp. 170–178, 2024, doi: 10.35870/jtik.v8i1.1416.

[22] M. Nurjannah, Hamdani, and I. F. Astuti, "Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) untuk Text Mining," *J. Inform. Mulawarman*, vol. 8, no. 3, pp. 110–113, 2013.

[23] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, pp. 640–651, 2021, doi: 10.30865/mib.v5i2.2937.

[24] M. Fadli and R. A. Saputra, "Klasifikasi Dan Evaluasi Performa Model Random Forest Untuk Prediksi Stroke," *JT J. Tek.*, vol. 12, no. 02, pp. 72–80, 2023, [Online]. Available: http://jurnal.umt.ac.id/index.php/jt/index

[25] F. K. Kartono, S. Nursaadah, M. R. Nugroho, and D. A. Tama, "Pengujian Black Box Testing Pada Sistem Website Osha Snack : Pendekatan Teknik Boundary Value Analysis," vol. 06, no. 02, pp. 754–766, 2024.