Jurnal Inovasi Teknologi dan Rekayasa
Vol. 9, No. 2, July-December 2024, pp. 527-535

Recommendation System for Selecting Electronic Products Using Content-Based Filtering Method

Exzaraja Indiandra Scandea [a,1,*], Rudi Susanto [a,2], Moh. Muhtarom [a,3]

[a] *Faculty of Computer Science, Informatics Engineering, Duta Bangsa University,* Indonesia
[1] *202030028@mhs.udb.ac.id*; [2] *rudi_susanto@udb.ac.id;* [3] *muhtarom@udb.ac.id*
* *corresponding author*

ARTICLE INFO

Article history:
Published

Keywords:
Content-Based Filtering,
Cousine Similarity,
Electronic Products,
Similarity,
Recommendation System

ABSTRACT

This study aims to develop a recommendation system for electronic products using the Content-Based Filtering method. In this research, calculations were performed to identify 12 products that exhibit similarity based on their respective categories. The method employed for these calculations was cosine similarity, which measures the degree of similarity between products. The analysis results indicate that there are three products with the highest similarity values, specifically 73, 71.15, and 64.89. These products were selected based on their relevant characteristic similarities and features, thereby providing accurate recommendations for consumers. This recommendation system is expected to significantly enhance the online shopping experience by assisting consumers in finding products that align closely with their preferences and needs. Consequently, this research offers substantial contributions to companies seeking to improve customer satisfaction and operational efficiency, ultimately leading to better consumer engagement and increased sales in the competitive electronic market.

I. Introduction

Current technological advances demand increased efficiency in various aspects of work, including in the product sales process [1]. Many companies, including CV. Cahaya Manunggal, still rely on manual systems that have proven to be less effective and efficient. Therefore, the implementation of a more modern computerized system becomes very important to replace outdated manual methods.

CV. Cahaya Manunggal, as one of the suppliers of electronic products in Sukoharjo City, still uses a manual system in its sales process. This causes the work process to be slow and inefficient, thus hindering growth and customer satisfaction. A system that is able to increase efficiency in sales and provide convenience for consumers in finding products that match their preferences is needed.

The purpose of this research is to design and implement a website-based electronic product recommendation system using the Content-Based Filtering method. This research aims to improve the efficiency of the sales process at CV. Cahaya Manunggal and assist consumers in choosing products that best suit their needs and preferences.

The implementation of the Content-Based Filtering algorithm will be done by analyzing available product data, such as descriptions, features, and categories [2]. This algorithm will utilize the information to provide relevant recommendations based on the similarity of product characteristics that users are interested in. For example, if users often choose products with certain features, the system will recommend other products that have similar features. This method not only improves the relevance of recommendations, but also allows users to explore a wider range of options according to their existing preferences.

The results of this research are expected to make a significant contribution to CV. Cahaya Manunggal in improving operational efficiency and customer satisfaction. In addition, this research can also be a reference for other companies that want to implement computerized systems and

DOI: 10.31572/inotera.Vol9.Iss2.2024.ID398 W : http://inotera.poltas.ac.id | E : inotera@poltas.ac.id

recommendation systems in their sales processes. Thus, this research focuses not only on system development, but also on improving user experience in online shopping.

II. Method

This research uses the content-based filtering method in the recommendation system is a method that considers user behavior from the past to identify patterns of behavior in order to recommend suitable items [3]. this method analyzes user preferences based on previous activities to build a model. the resulting model will be matched with a set of attribute characteristics of the items to be recommended. items with the highest match rate will be considered as recommendations for users [4].

A. Data Pre-processing

Pre-processing is a stage aimed at selecting and preparing raw data for further processing. This stage focuses on transforming data into an analyzable format, with the primary objective of representing each document as a feature vector [5]. The process involves the separation of words that comprise the documents and the application of specific techniques to enhance the quality of the resulting information. The selection of appropriate pre-processing methods is highly dependent on the characteristics of the existing data, which in turn can influence the accuracy and relevance of the final results of the analysis conducted [6].

B. Weighting

Weighting in recommendation systems often utilizes the vector space model method to model each document that represents items to be considered as recommendation candidates [7]. Each document is represented in the form of a matrix containing terms and their weight values, where each weight reflects the level of importance of the term in the context of the document in question. There are several weighting methods such as term frequency, inverse document frequency, and TF-IDF. Each document will first go through a pre-processing stage to prepare the raw data for weighting.

1. Term Frequency

Term frequency is one of the most basic weighting methods to give a weight value to a term. Each term is considered to have a level of importance in a document that is proportional to the number of occurrences in the document. This method utilizes the frequency of occurrence of terms in each document. The weight value for each term in each document is calculated based on the logarithm of the frequency of occurrence of the term. The formula for calculating term frequency can be seen in Equation 1.

$$w_{t,d} = \begin{cases} 1 + \log_{10} tf_{t,d},\ jika\ tf_{t,d} > 0 \\ 0,\ jika\ tf_{t,d} = 0 \end{cases} \tag{1}$$

Description:

- *ft,d* = the number of occurrences
- (t) = term
- (d) = document

Equation 1 above shows the number of occurrences of terms in a document. The use of logarithms aims to reduce significant differences in the frequency of occurrence between terms, so that they are not too large. Each term weight value with a frequency greater than zero will be added 1, to distinguish between term weight values that appear once and those that do not appear at all.

2. Inverse Document Frequency

The inverse document frequency weighting method considers that any term that rarely appears (not found in many documents) is considered to have a higher level of importance than a common term (often found in many documents). The weight value for each term is considered inversely proportional to the number of documents containing that term. This method utilizes the number of documents containing the term as well as the total number of documents as a whole. The weight value for each term is calculated by using the logarithm of the total number of documents divided by the number of documents containing the term. The formula for calculating inverse document frequency can be seen in Equation 2.

$$idf_t = \log_{10}(\frac{N}{df_t}) \tag{2}$$

Description:

- N = number of documents
- *dft* = number of documents containing term(t)

In Equation 2, dft, which indicates the number of documents containing term tt, will always have a value smaller or equal to NN, which is the total number of documents.

3. TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) is a statistical metric that describes the relevance of a term to a number of documents, so that the term can serve as a keyword for a particular document. Through this value, some documents can be identified or categorized more precisely. TF-IDF method is one of the widely used term weighting techniques. The TF-IDF weight is obtained from multiplying the tf (Term Frequency) value for a term in a particular document and the idf (Inverse Document Frequency) value of that term. The formula for calculating TF-IDF can be seen in Equation 3.

$$\mathrm{w}_{t,d} = \mathrm{w}_{\mathrm{tf}_{t,d}} \times idf_t \tag{3}$$

Description:

- Wtft,d = log TF weight of term (t) in document (d)
- *idft* = inverse document frequency value on term (t)

Normalization on term weighting can be seen in Equation 4:

$$\mathrm{w}_{t,d} = \frac{\mathrm{w}_{t,d}}{\sqrt{\sum_{t=1}^{n} \mathrm{w}_{t,d}{}^2}} \tag{4}$$

Description:

- Wt,d = TF-IDF weight of term (t) in document (d)
- n = number of terms

Equation 3 shows the weight of the term frequency value for term tt in document d, which is multiplied by the inverse document frequency value of the term. In the normalization described in Equation 4, the weights generated by TF-IDF will be converted into the range [0,1] by dividing them by the document length. This step aims to reduce significant differences between term weight values that have a large difference.

C. Similarity Calculation

Cosine similarity is a method used to measure the similarity between two distinct documents by calculating the cosine of the angle formed by the vectors representing each document [4]. The features within a document serve as dimensions that form a vector. The similarity between two documents can be assessed by calculating the distance between the resulting vectors. Various methods exist for calculating the distance between two vectors, including Euclidean distance and cosine similarity. Generally, the cosine similarity method is employed in data mining, information retrieval systems, and recommendation systems to evaluate the similarity between two document vectors. The formula for cosine similarity can be found in Equation 5, while the normalization of TF-IDF is presented in Equation 6.

$$\cos_{(q,d)} = \frac{q \times d}{|q| \times |d|} = \frac{\sum_{i=1}^{|v|} q_i d_i}{\sqrt{\sum_{i=1}^{|v|} q_i {d_i}^2 \times \sum_{i=1}^{|v|} q_i {d_i}^2}} \tag{5}$$

Description of Equation 5:

- q = TF-IDF weight of the query
- d = TF-IDF weight of the document

With normalization on term weighting:

$$\cos_{(q,d)} = q \times d = \sum_{i=1}^{|v|} q_i d_i \quad (6)$$

Description of Equation 6:

- q = TF-IDF weight of the query
- d = TF-IDF weight of the document

The results obtained from Equations 5 and 6 will be in the range [0,1]. The larger the resulting value (closer to 1), the smaller the angle formed by the two vectors, indicating that the two documents being compared are more similar. Conversely, the smaller the value obtained (closer to 0), the larger the angle formed by the two vectors, indicating that the two documents are more different. Thus, it can be concluded that the degree of similarity is directly proportional to the cosine value.

D. *Filtering by Category*

Category-based filtering is an approach used to filter and recommend products based on relevant category groupings. In this recommendation system, products are grouped into various categories that reflect their characteristics and features. For example, in the context of electronic products, categories may include televisions, laptops, smartphones, and other devices. By implementing a category-based filtering algorithm, the system can provide more focused recommendations to users based on the categories they have previously selected or browsed. This not only makes it easier for users to find products that match their preferences, but also improves the shopping experience by reducing the time needed to search for products [8]. Category-based filtering is thus an effective tool in improving recommendation relevance and customer satisfaction.

III. Results and Discussion

The process of selecting animal feed recommendations is carried out through several stages, including word processing, TF-IDF measurement, and similarity calculation. This research will calculate the number of products that will be recommended if the customer selects the keyword Category = “Lamp”, Description = “Energy-Efficient LED Lamp”, and Price = “50000”.

A. *Data Pre-processing*

Data Preprocessing is a step that involves selecting raw data to be processed for each document, including tokenization, text processing, filtering, and derivation. The primary objective of this step is to represent each document as a vector by separating the words that compose the document. Text preprocessing is performed on the name, brand, category, description, and price to enable the data to be transformed into a numerical format using TF-IDF and cosine similarity [9]. The stages in this process include cleaning, case folding [10], tokenization [11], lemmatization [12], and stop word removal [13]. Preprocessing is conducted on items that contain only keywords to ensure that the search process is efficient, utilizing product data consisting of 24 data. The results of the preprocessing can be seen in the following Table 1.

Table 1. Text Preprocessing Result Table

Data	Results
Blender Miyako	blender miyako blender 2in1 kapasitas 1 liter sni 238000
Blender Philips	blender philip kapasitas 2 liter 440000
Lampu Luby Prima	lampu luby prima 7watt low voltase led 82000
Lampu Luby Tricolour	lampu luby tricolour led rechargeable 9watt white yellow warm white 60000
Lampu Luby Diamond	lampu luby diamond 24watt led hemat energi 40000
Lampu Philips Esential	lampu philip esential 5watt led hemat energi sni 20000
Kipas Luby Minifan	kipas luby minifan mini hemat energi 115000

Kipas Luby Portablefan	kipas luby portablefan angin portable 145000
Setrika Miyako Ei1008M	setrika miyako ei1008m dengan permukaan tidak lengket memiliki konsumsi daya sebesar 395w 150000
Setrika Maspion EX1010	setrika maspion ex1010 dengan permukaan tidak lengket memiliki konsumsi daya sebesar 350w 170000
TV Samsung Crystal	tv samsung crystal smart 4k 43 inc uhd 4499000
TV LG Dynamic	tv lg dynamic 32 inc hd triple protection 1900000
Speaker Niko Beat	speaker niko beat active 35w support fm radio usb sd card bluetooth 350000
Speaker Advance Digitals Duo	speaker advance digitals duo komputer usb with volume control 80000
Kipas Sekai DFN 607	kipas sekai dfn 607 angin meja 25w dengan ukuran balingbaling 15cm 120000
Kipas Miyako KAD927B	kipas miyako kad927b angin meja dari material plastik konsumsi daya mencapai 30w 210000
Lampu Luby Classic	lampu luby classic led classic 10watt 90000
Lampu Luby Emergency	lampu luby emergency 12watt 120000
Lampu Luby Capsule	lampu luby capsule luby lampu lampu led 8wattt low voltase 80000
Lampu Luby Mercury	lampu luby mercury luby lampu lampu bohlam 50watt white warm yellow 100000
Lampu Philips Mini Bulb	lampu philip mini bulb philip lampu lampu 7watt hemat energi 24000
Lampu Philips LED Stick	lampu philip led stick 15watt ramah lingkungan material sintetis 54000
Lampu Philips LED Bright	lampu philip led bright philip lampu lampu 30watt high lumen led 40000
Lampu Philips Candle	lampu philip candle 8watt hemat energi white yellow warm white 32000

B. Weighting

1. Term Frequency

The results of the Term Frequency calculation can be seen in Table. 2, which is used to calculate the frequency of each term in the document. This process also applies a logarithmic transformation that produces vector values for each feature from 24 data, as follows:

Table 2. Term Frequency

Term	TF
blender	0,272727
miyako	0,181818
2in1	0,090909
kapasitas	0,090909
liter	0,090909
…	…
speaker	0,1875
niko	0,125
beat	0,0625
active	0,0625
35w	0,0625
…	…
hemat	0,071429
energi	0,071429
white	0,142857
yellow	0,071429
warm	0,071429

2. Inverse Document Frequency

The results of the Inverse Document Frequency (IDF) calculation, shown in Table 3, IDF is calculated based on the number of documents containing a particular term. This process also applies a logarithmic transformation, which results in vector values for 24 data as follows:

Table 3. IDF

Term	**IDF**
blender	3,178054
miyako	3,178054
2in1	3,178054
kapasitas	3,178054
liter	3,178054
…	…
speaker	3,178054
niko	3,178054
beat	3,178054
active	3,178054
35w	3,178054
…	…
hemat	3,178054
energi	3,178054
white	3,178054
yellow	3,178054
warm	3,178054

3. TF-IDF

After calculating TF and IDF, the next step is to calculate the vector for each data feature using the TF-IDF formula shown in Table 4. This process calculates the combination of TF and IDF vector values, resulting in the TF-IDF calculation for the features of the 24 data as follows:

Table 4. TF-IDF

Term	**TF-ITF**
blender	0,866742
miyako	0,577828
2in1	0,288914
kapasitas	0,288914
liter	0,288914
…	…
speaker	0,595885
niko	0,397257
beat	0,198628
active	0,198628
35w	0,198628
…	…
hemat	0,227004
energi	0,227004
white	0,454008
yellow	0,227004
warm	0,227004

C. Similarity Calculation

Previous research by [14] showed that the recommendation system studied, entitled 'Recommendation System Content-Based Filtering Men's Skincare at E-Commerce Shopee', produced a similarity value of 0.5611, which reflects the low accuracy in providing relevant recommendations to users. This value indicates that the system still has limitations in effectively capturing user preferences and needs. In this context, this research aims to update and improve the calculation of the recommendation system through the application of a more sophisticated Content-Based Filtering method. By integrating a more in-depth analysis of product characteristics and user preferences, it is expected that this research can achieve higher similarity values, as well as improve the relevance and satisfaction of users in receiving electronic product recommendations that match their needs. Furthermore, researchers implemented a content-based filtering-based recommendation system to provide electronic product recommendations using cosine similarity. From the analysis, it produces similarity values as in Table 5

Table 5. Similarity Calculation.

Product	Cosine Simmilarity
Blender Miyako	0
Blender Philips	0
Lampu Luby Prima	56,78
Lampu Luby Tricolour	50,52
Lampu Luby Diamond	73
Lampu Philips Esential	71,15
Kipas Luby Minifan	16,67
Kipas Luby Portablefan	0
Setrika Miyako Ei1008M	0
Setrika Maspion EX1010	0
TV Samsung Crystal	0
TV LG Dynamic	0
Speaker Niko Beat	0
Speaker Advance Digitals Duo	0
Kipas Sekai DFN 607	0
Kipas Miyako KAD927B	0
Lampu Luby Classic	55,34
Lampu Luby Emergency	48,67
Lampu Luby Capsule	56,78
Lampu Luby Mercury	47,43
Lampu Philips Mini Bulb	64,89
Lampu Philips LED Stick	54,01
Lampu Philips LED Bright	60,3
Lampu Philips Candle	57,74

D. *Filtering by Category*

These products were then analyzed by category to determine their similarity to other products in the dataset, which resulted in 12 products with the highest similarity value to the identified products. After applying the category-based filtering method, the system successfully identified and recommended a total of 12 products that matched the selected categories. These products include various relevant options, such as category, description, and price, which have been grouped based on their characteristics and features. Thus, users can easily find the options that best suit their preferences, thereby increasing the effectiveness and satisfaction of the shopping experience.

Table 6. Filtering by Category

Product	Cosine Simmilarity
Lampu Luby Diamond	73
Lampu Philips Esential	71,15
Lampu Philips Mini Bulb	64,89
Lampu Philips LED Bright	60,3
Lampu Philips Candle	57,74
Lampu Luby Prima	56,78
Lampu Luby Capsule	56,78
Lampu Luby Classic	55,34
Lampu Philips LED Stick	54,01
Lampu Luby Tricolour	50,52
Lampu Luby Emergency	48,67

Lampu Luby Mercury	47,43

IV. Conclusion

This research successfully developed an electronic product recommendation system using the Content-Based Filtering method, which calculates product similarity by category. In this process, the system analyzes product features, such as descriptions and specifications, to determine similarities between different products.

The calculation results show that the system can generate 12 products that have a high degree of similarity based on the same category. The three products with the highest similarity values have similarity 73, 71.15, and 64.89 respectively. These values reflect how close the characteristics of the product are to the product searched by the user. Thus, the system not only improves product search efficiency, but also provides relevant and satisfying recommendations for consumers, thus improving the overall shopping experience.

To improve the quality and effectiveness of the system, it is suggested that further development be carried out by considering other methods, such as collaborative filtering, to improve the accuracy of recommendations. Conducting system trials with real users is also important to get constructive feedback. In addition, more data collection related to user and product preferences would be beneficial to improve the performance of the algorithm. Integration of recommendation systems with other technologies, such as big data analytics and machine learning, can provide deeper insights into consumer behavior. Finally, educating users on the use of recommendation systems will help maximize their shopping experience.

Acknowledgment

I would like to thank all those involved in the academic and research process. They have given their time, energy, and thoughts to help and provide guidance that is very meaningful to researchers.

References

[1] H. Indrayani, "The application of information technology in increasing the effectiveness, efficiency and productivity of companies," *J. El-Riyasah*, vol. 3, no. 1, pp. 48–56, 2012.

[2] A. Syaifuddin and M. Ningsih, "Application of Content-Based Filtering Method in Marketing Communication Strategy on Tokopedia Marketplace," *J. Responsive Ris. Science and Inform.*, vol. 5, no. 2, pp. 185–194, 2023.

[3] R. Oktavian and F. Amin, "Designing a Laptop Recommendation System with a Prototyping Model and the Application of a Content-Based Filtering Approach at ELS Computer Shop Semarang," *G-Tech J. Teknol. Apply.*, vol. 8, no. 1, pp. 66–80, 2023.

[4] E. Salim, J. Pragantha, and D. L. Manatap, "Designing a Film Recommendation System Using the Content-based Filtering Method," *J. Pengemb. Technol. Inf. and Komput Science.*, vol. 5, no. 6, pp. 2188–2199, 2021.

[5] D. Hartanti, A. I. Pradana, and S. Lestari, "Comparative Algorithm of Decision Tree, SVM and ANN for Hotel Reservation," *DutaCom*, vol. 16, no. 1, pp. 21–27, 2023.

[6] I. Sholeh, N. Nurchim, and R. Susanto, "Detection of the Use of Number of Shelves Based on Image Processing on Shelf Layout Design Images," *J. Komput. and Technol.*, vol. 5, no. 1, pp. 1–10, 2024.

[7] D. P. Indini, K. Khairunnisa, N. D. Puspa, T. A. Siregar, M. Mesran, and M. Kom, "Application of OCRA Method in Determining the Best Online Learning Media during the Covid-19 Pandemic with ROC Weighting," *J. Sist. Comput. and Inform.*, vol. 3, no. 2, pp. 60–66, 2021.

[8] M. R. Alifi, M. Muhtarom, and I. Oktaviani, "Implementation of Information Systems with the Servqual Method for Web-Based Patient Satisfaction Surveys," in *Proceedings of the National Seminar on Information Technology and Business*, pp. 719–724, 2023.

[9] A. Setya, P. Widyaningsih, and S. Sumarlinda, "Animal Nutrition Selection Recommendation System with Content Based Filtering Method," vol. 9, no. 2, pp. 363–369, 2024.

[10] H. S. Putra, V. Atina, and D. Hartanti, "Application of Naive Bayes Algorithm for Sentiment Analysis

of Service and Facility Satisfaction (Case Study: PKU Muhammadiyah Sukoharjo General Hospital)," *J. Adv. Inf. Ind. Technol.*, vol. 6, no. 1, pp. 61–72, 2024.

[11] I. H. Junaidi and S. Sumarlinda, "Retrieval System of Tirta Asata E-Archives Using Vector Space Model Algorithm," *J. FASILKOM*, vol. 14, no. 2, pp. 361–369, 2024.

[12] Z. Abidin and A. Junaidi, "Text Stemming and Lemmatization of Regional Languages in Indonesia: A Systematic Literature Review," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 10, no. 2, pp. 217–231, 2024.

[13] A. Schofield, M. Magnusson, and D. Mimno, "Pulling out the stops: Rethinking stopword removal for topic models," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, short papers*, pp. 432–436, 2017.

[14] A. D. Saputro and F. Amin, "Sistem Rekomendasi Content-Based Filtering Skincare Pria Di E-Commerce Shopee," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 7, no. 1, pp. 106–113, 2024.