KOMPUTASI: JURNAL ILMIAH ILMU KOMPUTER DAN MATEMATIKA
Vol. 21 (1)(2024), 31-44 p-ISSN: 693-7554, e-ISSN:2654-3990

Recommender Systems Using Hybrid Demographic and Content-Based Filtering Methods for UMKM Products

Salsa Nadira Putri[1,*], **Tjut Awaliyah Zuraiyah**[2], **Dinar Munggaran Akhmad**[3]

[1]Computer Science, Faculty of Mathematics and Natural Science, Pakuan University, Bogor, West Java, 16143, Indonesia

Abstract

Marketing digitization such as e-commerce is needed by micro, small and medium enterprises (UMKM) in Bogor City and Regency so that the products are more easily accessible to consumers. One of the digital marketing that is commonly used by consumers is an e-commerce website. The Recommendation System is implemented into e-commerce websites to increase consumer convenience in online shopping. The recommendation systems method applied is Demographic Filtering and Content-based Filtering. Demographic Filtering uses IMDB Weighted Rating calculations which generate recommendations globally and give recommendations based on each product's IMDB Weighted score. Content-based Filtering uses Cosine Distance calculations which generate personal recommendations and give recommendations based on the score cosine distance of each product in the form of a presentation of the similarity of products that have been purchased with other products. This research uses 107 UMKM fashion and craft product data that was obtained from Bogor City Regional Craft Center which sells various kinds of UMKM products from Bogor City and Regency. Data preprocessing is then carried out on the raw data, with the Data Cleaning, Data Transforming and Data Splitting stages which divide the data in a ratio of 80:20. The accuracy of Demographic Filtering Recommendation System reaches 82.7% and Content-based Filtering Recommendation System reaches 100%.

Keywords: *UMKM; Recommender System; Demographic Filtering; Content-based Filtering*

1. Introduction

Micro, small, and medium enterprises (UMKMs) are business activities carried out by individuals or individuals, households, or small-scale business entities [1]. UMKMs as a support for the state and regional economy still have difficulties in marketing their products, so a special platform is needed to sell UMKM products [2].

Digital marketing is one type of marketing being widely used to promote products or services and to reach consumers using digital channels [3]. With digital marketing, transaction processes can be carried out at any time (real-time) and products can be marketed globally. Therefore, nowadays people are starting to abandon traditional marketing models little by little and are switching to modern marketing, namely digital marketing [2].

[*]**Corresponding author:** ***E-mail adress:*** *065122182@student.unpak.ac.id*
Received: 29 May 2023, Accepted: 11 July 2023 and available online 31 July 2023
https://doi.org/10.33751/komputasi.v20i2.7905

One example of implementing Digital Marketing that is widely used by the public is E-Commerce. E-commerce is the activity of distributing, selling, buying, and marketing products using telecommunications networks such as the Internet or computer networks. E-commerce as a field of online transactions is increasingly in demand by the public. With e-commerce, the process of buying and selling goods and services becomes cheaper, more practical, and more efficient [4].

E-commerce is usually available on various platforms ranging from social media, websites, to mobile applications. This research uses a website as an E-Commerce platform. A website is a collection of pages from a domain that contains various information so that it can be read and viewed by Internet users. With a website, various kinds of information become easier to disseminate [5].

Recommender system (RS) has become a very important factor in many E-Commerce sites. The typical Recommender Systems are software tools and techniques that provide support to people by identifying interesting products and services in online stores. It also provides a recommendation for certain users who search for recommendations [6]. Generally, recommender system methods can be classified as collaborative filtering, content-based filtering, demographic filtering, and knowledge-based filtering [7]. Demographic Filtering is since users have common personal attributes, such as gender, age, country, etc. These general personal attributes can be called demographic attributes. Based on this, the system can make recommendations by categorizing users using demographic attributes, thereby providing global recommendations. This approach is very useful, especially if the amount of information you have is small [8]. In general, there are 3 stages in the Demographic Filtering method, namely filtering, scoring, and sorting [9]. Content-based Filtering provides recommendations for items that have similar content features to items that the user liked before. The approach used uses item information to be recommended based on the user's profile so that recommendations are given personally [8]. Content-based filtering does not need rating data to generate item recommendations since it uses the product descriptions' similarity [9]. In general, there are 3 stages in the Content-based Filtering method, namely Encoding Content into Banks, Document Search and Sorting the data [10]. The combination of more than one method in a recommender system is frequently called hybrid filtering [11]. The concept of personalization is a fundamental requirement for e-commerce [8]. With the Recommendation System feature, consumers are expected to get product recommendations according to their needs and personalization.

The main requirement in the recommender system is the data [9]. Furthermore, data preprocessing is carried out on the data that has been obtained so that the data can be processed according to the recommendation system method used. Preparing data is a very important preprocessing stage in data mining, the main reason is that the quality of the input data greatly affects the quality of the resulting analysis output [12]. Raw data is usually susceptible to missing values, noisy data, incomplete data, inconsistent data, and outlier data. Preprocessing data is an essential step to enhance data efficiency. Data preprocessing steps deal with data preparation and transformation of the dataset to make knowledge discovery more efficient [13]. Preprocessing data in this research using data cleaning, data transforming and data splitting. Data cleaning is considered an effective approach to improving data quality to help practitioners and researchers be devoted to downstream analysis and decision-making without worrying about data trustworthiness [14]. Data cleaning is used for cleaning the unnecessary data, filling the null column, and grouping the data into categories so that the dataset can be used more efficiently and easily. Data transforming is used for the normalization and generalization of data so that the value of the dataset does not have too far range. Data normalization is one of the pre-processing approaches where the data is either scaled or transformed to make an equal contribution to each feature [15]. Data splitting is the act of partitioning available data into two portions, usually for cross-validatory purposes. One portion of the data is used to develop a predictive model and the other to evaluate the model's performance [16].

The validation test is carried out, the accuracy and validation score of the recommendation method is calculated by calculating the MAE (Mean Absolute Error) value for the Demographic Filtering recommendation system. MAE states the prediction error value in a model. The MAE value is the average value of the total absolute error value (the difference between the predicted value and the actual value [17]. Precision Recall is used for the Content-based Filtering recommen-

dation system. Precision and Recall values determine the number of similarities between sentences with a range of 0-1 [18].

2. Methods

The research method used in this study is the System Development Life Cycle (SDLC) approach, the SDLC Flowchart can be seen in Figure 1.

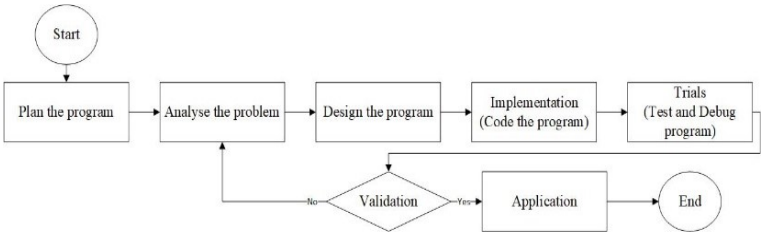

Figure 1. SDLC Diagram

2.1. Plan the Program

This stage is carried out to study information and data needs in the recommendation system. This stage is carried out by analyzing the system that has been made before and looking for weaknesses and strengths so that the results of the analysis can be applied to the system being developed.

2.2. Analyze the Problem

The overall Analyze the Problem stage is divided into the following 2 stages. The first stage is Observation, Interview and Study of Literature. This stage is carried out to find the data needed by the system and research references. The second stage is Data Mining (Data Collecting) and Data Processing. At this stage, the researcher visited the Bogor City Handicraft Center as a reference place for UMKMs to collect various data and information needed by the system. Furthermore, the raw data that has been collected is analyzed and processed so that the data becomes ready for use using the data preprocessing method including Data Cleaning, Data Transforming and Data Splitting which divides test data and training data into 80:20.

2.3. Design the Program

The overall system design stage is divided into 3 stages. The first stage is database design based on database needs, using ERD (Entity Relationship Diagram) and table specifications. The second stage is general system design describing how the system works in general by using a flowchart. The last stage is detailed system design by describing how the system works in detail using a flowchart. The General System Design Flowchart can be seen in Figure 2.

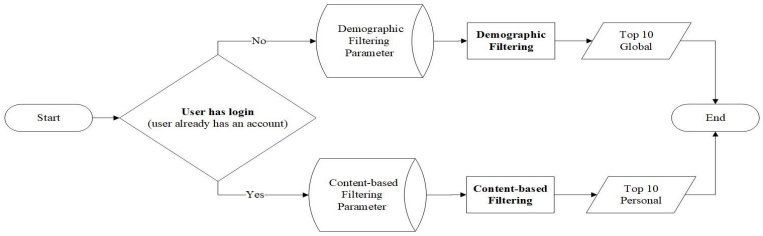

Figure 2. General System Design

2.4. Implementation (Code the Program)

This stage is the stage of making the system (coding). The recommendation system is made using the Hybrid Demographic Recommendation System and Content-based Filtering method. The Demographic Filtering method is used for users who have not logged in, this method is used to provide overall (global) recommendations to users. The calculation used for this method is the IMDB Weight Rating. This calculation is carried out to find the score value of each product to be sorted and used as a recommendation. The Content-based Filtering method is used for users who are logged in and have purchased a product, this method is used to provide personal recommendations to users. The calculation for this method is the Cosine Distance which is used to calculate the similarity of words between products. The system was built using the programming languages PHP, HTML, JavaScript and CSS and using the Laravel 10 framework. For the database part of the system, it uses MySQL and XAMPP as web servers and Python to analyze data. For the text editor program using Microsoft Visual Studio Code. At this stage, the appropriate design will be applied to the system.

2.5. Trials (Test and Debug Program)

The purpose of this stage is to evaluate the system whether the system that has been made is by the actual one, and whether there is an error, damage or failure in the system. There are 2 stages of the trial stage. The first stage is the structural test. This stage aims to find out whether the system is well structured, emphasizing application features that meet user needs, and whether it is good and correct by the design that has been made. The second stage is the functional test. The purpose of this stage is to find out the navigation and validation process, and whether the application created can function well or not according to its function.

2.6. Validation

The purpose of this stage is to find out whether the results of a system match the expected results. A validation test is a test to check the accuracy of the data that has been entered (input) into the application and whether the resulting (output) is by the design that has been made or not. After the validation test is carried out, the accuracy and validation score of the recommendation method is calculated by calculating the MAE (Mean Absolute Error) value for the Demographic Filtering recommendation system. Calculating Precision Recall for the Content-based Filtering recommendation system.

2.7. Application (Implementation)

This stage is a stage in the form of the final goal of the system, namely an E-Commerce recommendation system for UMKM products which in the future can be used by the public and sellers of UMKM products.

3. Result and Discussion

3.1. Data Preparation

a. Data Collecting

In total, there were 107 UMKM fashion and craft product data obtained by visiting the reference research site directly, namely the Bogor City Regional Craft Center which sells various kinds of UMKM products from Bogor City and Regency. The data obtained are in the form of product images, product prices, and tables of interview results which contain the information needed to support the research.

b. Observation, Interview and Study of Literature

After observing various E-commerce websites and existing marketplaces as examples and references for researchers. Interviews were conducted to find the required method parameters. The results of the interview are in the form of product criteria data with criteria 1-6 which are categorized based on sales of the product type, from those that sell the fastest, to those that take the longest to sell. The interviewed results that contain product criteria can be seen in Table 1.

Table 1. Product Criteria

Order of Product Criteria (1-6)	Product Criteria
1	Accessories
2	Clothes
3	Cloth
4	Bag
5	Toy
6	Decoration

3.2. Data Preprocessing

a. Data Cleaning

Based on the parameters obtained, the following data table is made, and the Data Cleaning process is carried out. In this step category and criteria column is added. The results can be seen in Table 2.

Table 2. Data Cleaning Results

id	**product name**	**price**	**price range**	**category**	**criteria**	**rating**	**total reviews**
1	Asbak Tembikar	100000	low	decor	6	3	1
2	Baju Batik Istana Bogor	325000	medium	fashion	2	5	4
3	Mainan Balok Pelangi	50000	low	toy	5	4	1
4	Bantal Twist Rope	250000	low	decor	6	3	1
5	Kain Batik Geometri	250000	low	fashion	3	5	3
...							
107	Wadah Tisu Simpel	40000	low	decor	6	3	1

b. Data Transforming

After the data cleaning process, the data is normalized so that the data is general and does not have a range that is too far. Here are the normalization equation using min-max normalization where X* is the attribute value after the normalization process, X is the attribute value before the normalization process, min(X) is the minimum value of attribute X and max(X) is the maximum value of attribute X.

$$X^* = \frac{X - min(X)}{max(X) - min(x)} \tag{1}$$

Here are the normalization parameters

Table 3. Normalization Parameters

Parameter	Minimum Value	Maximum Value
x1	1	3
x2	1	6
x3	1	5
x4	1	5

c. Data Splitting

Data that has been normalized is then divided into data train and data test with a ratio of 80:20.

Table 4. Data Train (80%)

No	x1	x2	x3	x4
1	0	0.8	0.75	0
2	0	0	1	1
3	0.5	0	1	1
4	0	0	1	1
5	0	0.6	1	0.333
85	0	0	0.5	0

Table 5. Data Test (20%)

No	x1	x2	x3	x4
1	0	0.6	1	0.333
2	0	1	0.5	0
3	1	0.6	0.5	0.333
4	0	1	0.5	0
5	0.5	0.4	0.75	0.667
22	0	0.6	1	0.333

3.3. Database Design

a. ERD (Entity Relationship Diagram)

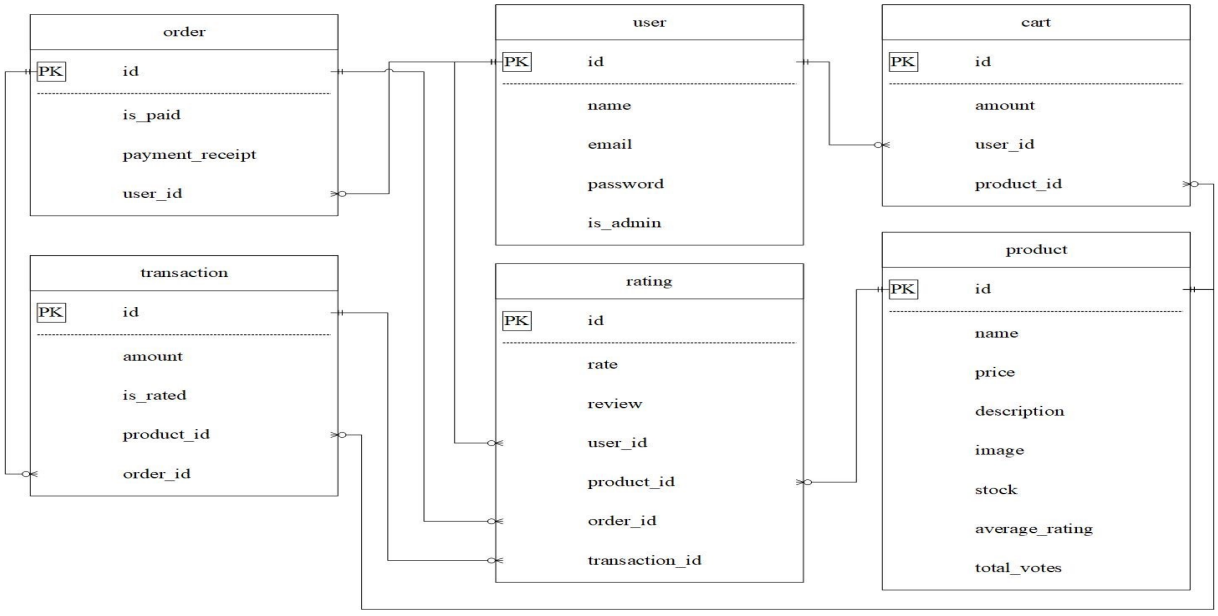

Figure 3. ERD

b. Table Specifications

In total, there are 6 tables used in this system. Among them are User Tables, Product Tables, Cart Tables, Order Tables, Transaction Tables and Rating Tables. Following are the specifications of the 2 main tables used in this system

Table 6. Product Table

Column	Data Type	Information
id	BigInt (20)	Primary Key (Not Null)
name	Varchar (255)	Not Null
price	Int (11)	Not Null
description	Longtext	Not Null
image	Varchar (255)	Not Null
stock	Int (11)	Not Null
average_rating	Double (100, 1)	Not Null
total_votes	Int (11)	Not Null

Table 7. Rating Table

Column	Data Type	Information
id	Bigint (20)	Primary Key (Not Null)
rate	Int (11)	Nullable
review	Longtext	Nullable
user_id	Bigint (20)	Foreign Key (Not Null)
product_id	Bigint (20)	Foreign Key (Not Null)
order_id	Bigint (20)	Foreign Key (Not Null)
transaction_id	Bigint (20)	Foreign Key (Not Null)

3.4. Demographic Filtering

The flow of the Demographic Filtering Recommendation System can be seen below.

a. Filtering data based on ratings and votes
   Sorting and defining the variables needed to process data, such as ratings, number of votes, quantile values, and the overall average rating.]

b. Scoring using the IMDB Weighted Rating
   Calculating the score value for each product using the IMDB Weight Rating formula. Below is the equation where v is the number of votes, m is the minimum requirement for data criteria, R is the average rating and C is the total average rating.

$$WR\frac{vR + mC}{v + m} \tag{2}$$

c. Sorting the scoring values
   Sort the score values that have been obtained from the largest to the smallest. Sorting results from the largest to the smallest score value of 107 product data. Table 8 shows the top 15 products based on IMDB Score after sorting.

Table 8. Top 15 products based on IMDB Score

ID	Product Name	R	v	m	C	Score
66	Piring Daun Talas	5	4	0.9	4.734	4.951
94	Teko Butterfly 1 Set	5	4	0.9	4.734	4.951
19	Kain Batik Paniisan	5	3	0.9	4.734	4.938
40	Kain Rayon Bambu	5	3	0.9	4.734	4.938
71	Selop Crochet	5	3	0.9	4.734	4.938
1	Asbak Tembikar	5	1	0.9	4.734	4.874

The IMDB Weighted Rating equation is then entered into the system and produces a Demographic Filtering Recommendation System which provides global recommendations intended for users who have not logged in (guests) and users who have logged in but have never ordered a product before.

The implementation of the Demographic Filtering Recommendation System on the product index page of the Bogor UMKM E-commerce website can be seen in Figure 4.

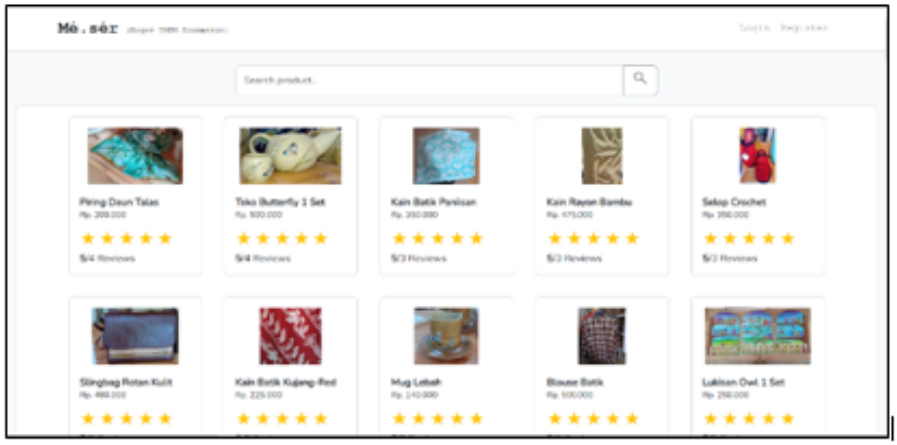

Figure 4. Demographic Filtering Implementation

3.5. Content-based Filtering

The flow of the Content-based Filtering Recommendation System can be seen below.

a. Preprocessing parameters data by Case Folding
   Change words from capital letters to lowercase. By using the strtolower function in PHP, all capital letters change to lowercase.

b. Encode content into a bank
   Separating words using Tokenizer and calculating the number of similar words using CountVectorizer. For example, if a user has purchased a product with product_id 0 and product_name "Batik Kujang Red Cloth", then the following product name data is known as a comparison attribute.

Table 9. Cosine Distance Attributes

product_id	product_name
0	Kain Batik Kujang Red
1	Kain Batik Bambu
2	Kain Batik Geometri
3	Baju Batik Kujang
4	Kain Batik Abstrak
5	Kain Batik Kujang

After Tokenizer and Countervectorizer, the Encoding results can be seen in Figure below.

	abstrak	baju	bambu	batik	geometri	kain	kujang	red
0	0	0	0	1	0	1	1	1
1	0	0	1	1	0	1	0	0
2	0	0	0	1	1	1	0	0
3	0	1	0	1	0	0	1	0
4	1	0	0	1	0	1	0	0
5	0	0	0	1	0	1	1	0

Figure 5. Demographic Filtering Implementation

From the table above, you can see the number and location of each word in each sentence, which is as follows.
D0: 0, 0, 0, 1, 0, 1, 1, 1
D1: 0, 0, 1, 1, 0, 1, 0, 0
D2: 0, 0, 0, 1, 1, 1, 0, 0
D3: 0, 1, 0, 1, 0, 1, 0, 0
D4: 1, 0, 0, 1, 0, 1, 0, 0
D5: 0, 0, 0, 1, 0, 1, 1, 0

c. Document Search
Find the percentage of word similarity between sentences using cosine distance calculations the equation can be seen below.

$$\text{Distance} = \cos(\theta) = 1 - \frac{\mu \cdot \nu}{\|\mu\|\|\nu\|} = 1 - \frac{\sum_{i=1}^{n} \mu_i \nu_i}{\sqrt{\sum_{i=1}^{n} \mu_i^2} \sqrt{\sum_{i=1}^{n} \nu_i^2}} \tag{3}$$

Based on calculations using the Cosine Distance equation, it is known that the percentage of words similar to D_0 compared to D_0 is 100%, D_1 is 57.7%, D_2 is 57.7%, D_3 is 57.7%, D_4 is 57.7%, and D_5 is 86%.

d. Sorting the word similarity percentage
Sort data that has the largest percentage of word similarity to the smallest. Based on the calculations above, it can be concluded that the order of product recommendations is the product with the 0th index (Kujang Red Batik Cloth), the product with the 5th index (Kujang Batik Cloth), the product with the 2nd index (Geometric Batik Cloth) or 3rd (Kujang Batik Shirt) or 4th (Abstract Batik Cloth) because it has the exact same percentage value. The Cosine Distance equation is then entered into the system and produces a Content-based Filtering Recommendation System which provides personal recommendations intended for users who have logged in and users who have logged in and have ever ordered a product before. The implementation of the Content-based Filtering Recommendation System on the product index page of the Bogor UMKM E-commerce website can be seen in Figure 6.

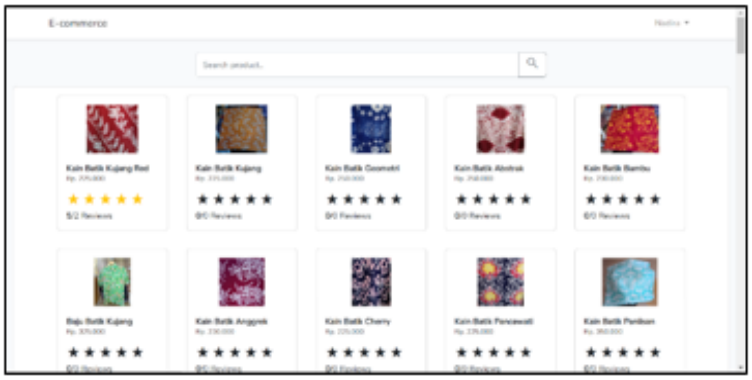

Figure 6. Content-based Filtering Implementation

3.6. Testing

a. Structural Testing

Compare the design of pages and forms with the implementation in the system. The following is the design and implementation of several main pages and forms.

Figure 7. Main Page Design

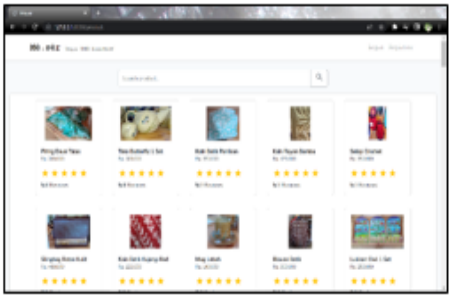

Figure 8. Main page Implementation

Figure 9. Create Product Page Form Design

Figure 10. Create Product Page Implementation

Figure 11. Show Product Page Design

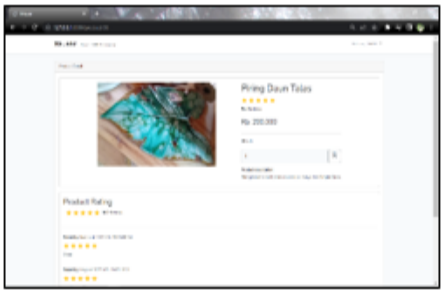

Figure 12. Show Product Page Implementation

b. Functional Testing

Test the functionality of the buttons on the page, the result of functional testing of several main features can be seen in Table 10.

Table 10. Top 15 products based on IMDB Score

Page	Button	Function	Result
Main Page	Login	Go to the Login page	functionate
Main Page	Register	Go to the Register page	functionate
Main Page	Home	Go to the Home page	functionate
Main Page	Search	Search for the product name	functionate
Halaman Login	Login	Go to the System	functionate
Halaman Register	Register	Create a new account	functionate
Create Product Page	Submit Data	Sending data to the database	functionate
Show Product Page	Add to Cart	Adding product to the cart	functionate
Show Order Page	Submit Payment	Sending proof of payment to the database	functionate
Order Page	Submit Rating	Send rating data to the database	functionate

c. Validation Testing

Compares the suitability of input and output pages, such as validating the input output of the following create product page.

Figure 13. Input: Create Product Pagen

Figure 14. Show Product Page

Based on the input and output above, it can be seen that the create product form page has successfully added product data to the database and displayed it on the show product page.

3.7. Validation

a. Demographic Filtering
Validation of this method uses MAE (Mean Absolute Error). Here's the equation.

$$MAE_i = \frac{1}{n}\sum_{i \in E}^{n} |p_i - r_i| \tag{4}$$

After the MAE value is obtained, the accuracy calculation is carried out using the formula below, with the X value being the maximum value of MAE.

$$\text{Accuracy} = \frac{x - \text{MAE}}{x} \times 100 \tag{5}$$

Based on the above equation, the MAE Demographic Filtering value is 0.17314 with an accuracy of 82.686%.

b. Content-based Filtering
Validation of this method uses Precision and Recall. Here's the equation.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{6}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \tag{7}$$

These two are obtained from calculating the number of True Positives (similar sentences), True Negatives (different sentences), False Positives (similar sentences that are incorrectly identified as different sentences) and False Negatives (different sentences that are identified as similar sentences). After the Precision and Recall values are obtained, accuracy calculations are carried out using the equation below.

$$Accuracy = \frac{TruePositive + TrueNegative}{Alldata} \tag{8}$$

Based on the above equation, the Precision and Recall Content-based Filtering values are each equal to 1 with an accuracy of 100

4. Conclusion

This research succeeded in providing recommendations in the form of global recommendations using a Demographic Filtering recommendation system based on the number of product ratings and reviews, as well as personal recommendations using a Content-based Filtering recommendation system based on the names of products that have been purchased by users. The recommendation system was successfully implemented on 107 Bogor MSME fashion and craft products. The Demographic Filtering method uses the MAE evaluation method to calculate the average absolute error. The MAE value obtained was 0.17314 with an accuracy value of 82.686%. Meanwhile, for the Content-based Filtering method, the evaluation method used is Precision Recall with the Precision and Recall values obtained each being 1 with an accuracy value of 100%.

5. Acknowledgement

Based on the research above after going through many tests and trials, it can be concluded that this system method has several shortcomings

Firstly, the rating used in this research is based on product quality obtained from the number of product sales based on the criteria and product price range. This is done because there is no historical rating data or data showing the quality of each product in any product marketing media. This can be overcome by looking for product data that already has rating attribute values and the

number of reviews for each product.

Secondly, in the Demographic Filtering method, if the admin adds new product data, the new product which has a rating of 0 is placed before the product which has a rating value of more than 0. This can happen because the IMDB Weight Rating score value of the product which has a rating of 0 is greater than the product who have a rating of more than 0. The IMDB Weight Rating score value is influenced by the average overall rating value (can be seen in equation (2)), so cases like this can occur. To prevent this, the system must first add a supporting recommendation system method using other variables, not just based on ratings and number of ratings so that the resulting recommendations are more precise and accurate.

Third, in the Content-based Filtering method, the recommendations displayed are only based on the names of products that the user has purchased. In the future, it is hoped that there will be additional methods so that users who have logged in and ordered will get recommendations based on ratings and names of products they have purchased so that the resulting recommendations will be better.

References

[1] T. Sudrartono et al., Kewirausahaan Umkm Di Era Digital. 2022.

[2] B. Susanto, A. Hadianto, F. N. Chariri, M. Rochman, M. M. Syaukani, and A. A. Daniswara, "Penggunaan Digital Marketing untuk Memperluas Pasar dan Meningkatkan Daya Saing UMKM," Community Empowerment, vol. 6, no. 1, 2020, doi: 10.31603/ce.4244.

[3] A. Yasmin, S. Tasneem, and K. Fatema, "Effectiveness of Digital Marketing in the Challenging Age: An Empirical Study," The International Journal of Management Science and Business Administration, vol. 1, no. 5, 2015, doi: 10.18775/ijmsba.1849-5664-5419.2014.15.1006.

[4] Akbar, N. Diana, and Afifudin, "Analisis Kinerja Keuangan Perusahaan Sebelum Dan Sesudah Penerapan E-Commerce (Studi Pada Perusahaan Sub Sector Retail Trade Dalam Index Saham Syariah Indonesia (Issi) Di Bursa Efek Indonesia)," Jurnal Ilmiah Riset Akuntansi, vol. 10, no. 03, pp. 23–32, 2021.

[5] M. Septiani, R. Aulianita, V. Sofica, and N. Hasan, "Sistem Informasi Penjualan Kayu Kusen Berbasis Website," Bianglala Informatika, vol. 9, no. 2, 2021, doi: 10.31294/bi.v9i2.11603.

[6] P. Melville and V. Sindhwani, "Recommender Systems 1 Definition," Ime.Usp.Br, pp. 1–21, 2017, [Online]. Available: https://www.ime.usp.br/ jstern/miscellanea/seminario/Melville1.pdf

[7] E. Çano and M. Morisio, "Hybrid recommender systems: A systematic literature review," Intelligent Data Analysis, vol. 21, no. 6. 2017. doi: 10.3233/IDA-163209.

[8] J. Chicaiza and P. Valdiviezo-Diaz, "A comprehensive survey of knowledge graph-based recommender systems: Technologies, development, and contributions," Information (Switzerland), vol. 12, no. 6, Jun. 2021, doi: 10.3390/info12060232.

[9] M. Johari and A. Laksito, "The Hybrid Recommender System of the Indonesian Online Market Products using IMDb weight rating and TF-IDF," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 5, no. 5, 2021, doi: 10.29207/resti.v5i5.3486.

[10] D. Daniel, B. Mulyawan, and T. Sutrisno, "PEMBUATAN APLIKASI E-COMMERCE BERBASIS WEB DENGAN FITUR REKOMENDASI MENGGUNAKAN METODE CONTENT-BASED FILTERING," Jurnal Ilmu Komputer dan Sistem Informasi, vol. 10, no. 1, 2022, doi: 10.24912/jiksi.v10i1.17837.

[11] X. Zhao, "A study on E-commerce recommender system based on big data," in 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analytics, ICCCBDA 2019, 2019. doi: 10.1109/ICCCBDA.2019.8725694.

[12] H. Junaedi, H. Budianto, I. Maryati, and Y. Melani, "Data Transformation pada Data Mining," Prosiding Konferensi Nasional Inovasi dalam Desain dan Teknologi, vol. 7, pp. 93–99, 2011, [Online]. Available: https://ideatech.stts.edu/proceeding2011/12-000113_INF_Hartarto_p93-99.pdf

[13] S. A. Alasadi and W. S. Bhaya, "Review of data preprocessing techniques in data mining," Journal of Engineering and Applied Sciences, vol. 12, no. 16, 2017, doi: 10.3923/jeasci.2017.4102.4107.

[14] X. Ding, H. Wang, G. Li, H. Li, Y. Li, and Y. Liu, "IoT data cleaning techniques: A survey," Intelligent and Converged Networks, vol. 3, no. 4, pp. 325–339, Dec. 2022, doi: 10.23919/ICN.2022.0026.

[15] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," Appl Soft Comput, vol. 97, Dec. 2020, doi: 10.1016/j.asoc.2019.105524.

[16] R. R. Picard and K. N. Berk, "Data splitting," American Statistician, vol. 44, no. 2, pp. 140–147, 1990, doi: 10.1080/00031305.1990.10475704.

[17] M. Naufal et al., "Implementasi Sistem Rekomendasi Makanan pada Aplikasi EatAja Menggunakan Algoritma Collaborative Filtering," Multinetics, vol. 7, no. 2, 2021.

[18] R. Singh and S. Singh, "Text Similarity Measures in News Articles by Vector Space Model Using NLP," Journal of The Institution of Engineers (India): Series B, vol. 102, no. 2, 2021, doi: 10.1007/s40031-020-00501-5.