Collabits Journal
Vol. 2 No. 3 September 2025 : 139-142
E-ISSN : 1979-5254, P-ISSN : 3062-8601
https://publikasi.mercubuana.ac.id/index.php/collabits

Improving E-commerce Platforms with Collaborative Filtering algorithms for Product Recommendations

Siti Maesaroh[1], Putri Nabila[2], Faiz Muhammad Ramadhan[3*]

[1,2,3] Computer Science study program, Universitas Mercu Buana, Indonesia

*Coressponden Author: faiz.m3110@gmail.com

Abstract - Online product reviews play a major role in the success or failure of an e-commerce business. In a transaction, buyers will usually find out information on the use of the product or service from online reviews posted by previous customers to get detailed product recommendations and make purchase decisions. Many reviews are created by users who often include strong sentimental opinions. This review of data is very promising and can be used by both customers and the Company. Customers can read reviews to know more about the quality of a product. However, due to the large number of reviews, it is difficult to see and read all consumer evaluations personally to get useful information. One effective approach in providing such recommendations is using Collaborative Filtering (CF) algorithms. This research aims to improve e-commerce platforms by applying Collaborative Filtering algorithms to provide more accurate and relevant product recommendations to users.

Keywords :
Algoritma Collaborative Filtering;
Product Recommendations;
New Heuristic Similarity;

Article History:
Received: 29-07-2025
Revised: 25-08-2025
Accepted: 15-09-2025

Article DOI : 10.22441/collabits.v2i3.27299

1. INTRODUCTION

In today's digital age, the e-commerce industry has become one of the fast-growing sectors, One way to enhance the experience is through the use of effective product recommendation systems. Product recommendations have become one of the key strategies in maintaining customer engagement and increasing sales within e-commerce platforms. By understanding buyer preferences and behavior, e-commerce platforms can provide recommendations that are relevant and according to user needs. This research focuses on a comparison between two potential algorithmic approaches, namely Collaborative Filtering and New Heuristic Similarity Algorithm, in the context of improving e-commerce platforms for product recommendations. The method in collaborative filering is Item-based-collaborative filtering is a recommendation method based on the similarity between rating a product and the product purchased. The recommendation system is a system that aims to estimate information that is of interest to its users and also helps potential customers in deciding what items to buy. Products that have the highest usability value are then used as recommendations. In this context, the two algorithms will be evaluated and compared in terms of performance, ease of implementation, and level of user satisfaction. Online reviews are becoming an important factor that encourages consumers to buy goods in e-commerce. In e-commerce, previous customer reviews can help buyers make better decisions by providing information about product quality, strengths and weaknesses, seller behavior, pricing, and delivery times. However, the existence of fake reviews poses a challenge in correctly assessing the sentiments expressed by genuine customers. Overall, the study aims to contribute to a further understanding of the use of algorithms in the context of e-commerce as well as offer practical guidance for developers and owners of e-commerce platforms in choosing the most suitable approach to improve the online shopping experience.

2. RESEARCH METHODOLOGY

2.1 literatur review

1. Study conducted by Ali Arifin. entitled "Application of Collaborative Filtering Algorithm System for Recommendation of Boarding House Selection Based on Rating" This study focuses on existing methods of collaborative filtering and item-based-collaborative filtering is a recommendation method based on similarities between rating a product and the product purchased. [1].

2. Another study conducted by Ahmad Syaiffudin. Titled "SISTEM REKOMENDASI PRODUK BERBAHASA INDONESIA PADA MARKETPLACE TOKOPEDIA MENGGUNAKAN METODE CONTENT BASE FILTERING" This study discusses the application of the Content Base Filtering method in the marketplace to provide product recommendations in Indonesian when users choose a product. [2].

3. Study conducted by Nuraini Siregar and samsudin "Implementation of Collaborative Filtering Algorithms in Mobile Based Food Menu Ordering and Recommendation Systems" This research focuses on the implementation of Collaborative Filtering algorithms for food ordering menus on mobile apps and recommendation systems [3].

4. Then the research conducted by Wang Juan. with the title " Survey of Recommendation Based on Collaborative Filtering" The discussion in this study is to discuss the main problems of the Collaborative Filtering algorithm, including data scarcity, cold start, and accuracy of similarity measurements. Furthermore, future research and development trends in integrating deep learning into recommendation systems are demonstrated. [4].

5. Then, research was conducted by Dita Aisha. With the title " SISTEM REKOMENDASI TOKO ONLINE MENGGUNAKAN ALGORITMA COLLABORATIVE FILTERING DAN CONTENT BASED FILTERING" This study is to compare and find out the difference between collaborative filtering algorithms and content based filtering in finding product recommendations [5].

2.2 Theoretical Basis

1. Shorten information and time

Basically, often a website or e-commerce application stores a lot of information in the form of product data or others. Recommendation systems are generally aimed at individuals who lack experience or competence which is enough to evaluate the number of alternative items that exist in a particular case. Simply put, it is a way to narrow information overload so that in a short time individuals will get a "conclusion" from a set of existing items (Rokhim &; Saikhu., 2016).

2. Costumer loyalty

A customer or user in an e-commerce needs a service that can serve well to get a shopping satisfaction. Customer Loyalty is to increase customer loyalty to online stores. With a recommendation system, all customer activities will be stored and used to provide special services for each customer so that customers are more happy to come and shop at online stores (Sari &; Sary, 2017).

3. Filtering

Users of e-commerce applications or websites sometimes need a filter for a product that is often related or items that are often searched. In the research of B.Thorat et al entitled survey on collaborative Filtering, content-based Filtering and hybrid recommendation system it was explained that content-based filtering algorithms try to recommend items based on the number of similarities. The most suitable items are recommended by comparing various candidate items with items previously rated by the user (B.Thorat et al., 2015).

3. RESULTS AND DISCUSSION

1. Methodologists

The dataset used in this study is the "Retailrocket Recommender System Dataset" downloaded from Kaggle. This dataset includes three main files: events.csv, item_properties.csv, and category_tree.csv. Data is processed and analyzed using Python with pandas, matplotlib, and seaborn libraries.

2. Data Exploration and Visualization

A.) Dataset

	timestamp	visitorid	event	itemid	transactionid
0	1433221332117	257597	view	355908	NaN
1	1433224214164	992329	view	248676	NaN
2	1433221999827	111016	view	318965	NaN
3	1433221955914	483717	view	253185	NaN
4	1433221337106	951259	view	367447	NaN

	timestamp	visitorid	event	itemid	transactionid
2756096	1438398785939	591435	view	261427	NaN
2756097	1438399813142	762376	view	115946	NaN
2756098	1438397820527	1251746	view	78144	NaN
2756099	1438398530703	1184451	view	283392	NaN
2756100	1438400163914	199536	view	152913	NaN

Figure 1. Dataset

B.) The distribution of event types (view, addtocart, transaction) in the dataset is shown in Figure 2.

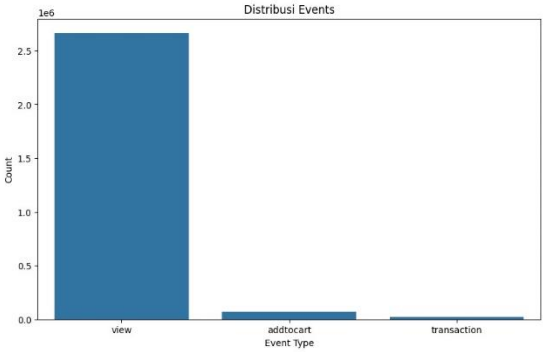

Figure 2. Visualisation of event distribution

Figure 2 shows that the most common event is 'view', followed by 'addtocart' and 'transaction'. This shows that users view products more often than they make a purchase or add them to their cart.

C.) Distribution of Events per Time

The distribution of events by time is shown in Figure 3.

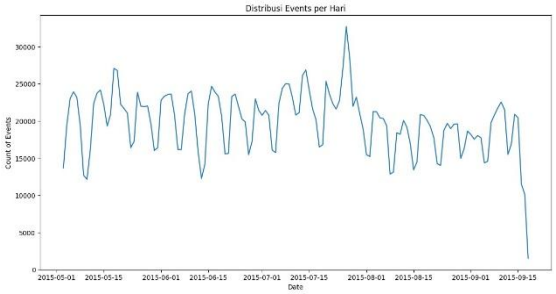

Figure 3. Visualisation of the distribution of events day

Figure 3 shows fluctuations in the number of events per day, which can provide insight into user behavior

over a period of time. The results of data exploration show that the majority of user interaction with e-commerce platforms is in the form of viewing products. This is consistent with user behavior that tends to do product searches and explorations before making a purchase decision. The distribution of events per day indicates the presence of seasonal patterns or trends that can be further analyzed to optimize marketing and sales strategies.

D.) Item category distribution

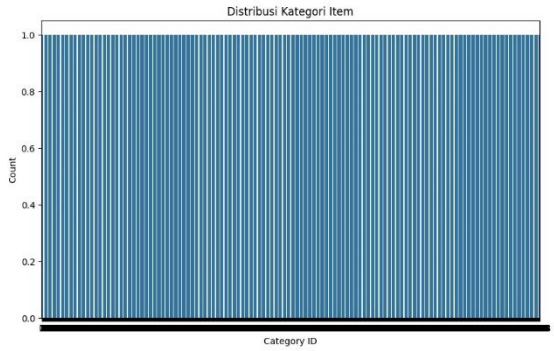

Figure 4. Visualisation of item category distribution

Visualization of item category distribution aims to show the distribution or number of items contained in each category. This helps us understand how diverse or popular a particular category is in the dataset.

E.) Heatmap correlation between item properties

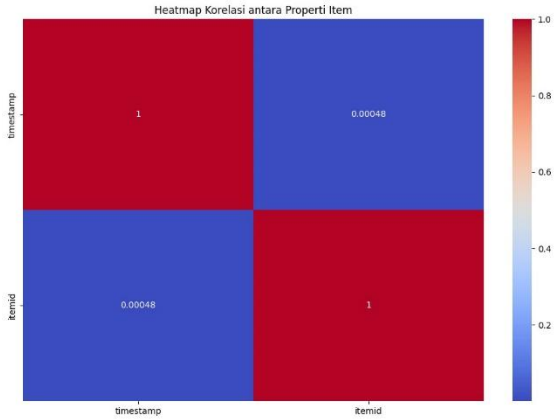

Figure 5. Visualisation of the correlation heatmap between item properties

The correlation heatmap between item properties aims to show the relationship or correlation between various attributes or properties owned by items in the dataset. This helps us understand whether there is a significant relationship between the properties of one item and another.

F.) Comparison of original ratings and predictions

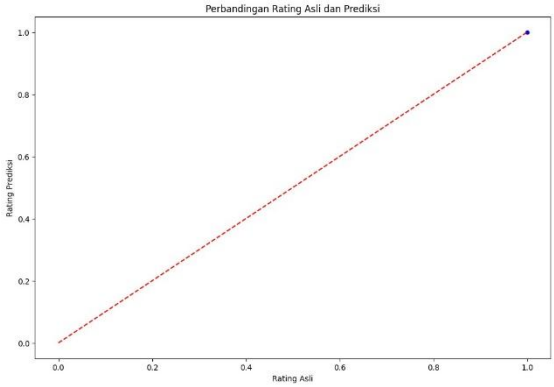

Figure 6. Visualisation of actual ratings and predictions

This image shows that some categories have a significantly larger number of items than others. This may indicate a particular focus of the e-commerce platform towards certain categories.

G.) Distribution of errors between original and predicted ratings

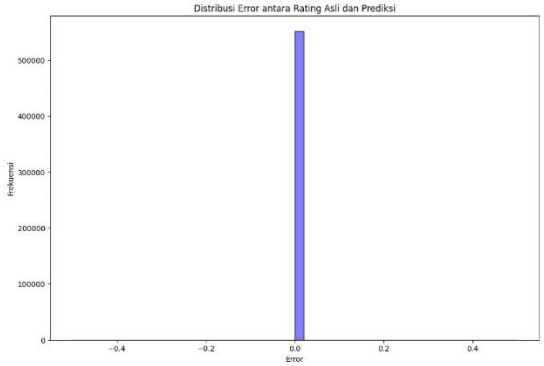

Figure 7. Error distribution results between original ratings and predictions

Figure 7 shows the distribution of errors between the original rating and the predicted rating. This histogram shows how far the model's prediction is from the actual value, where smaller errors indicate more accurate predictions.

H. Top 10 most popular items

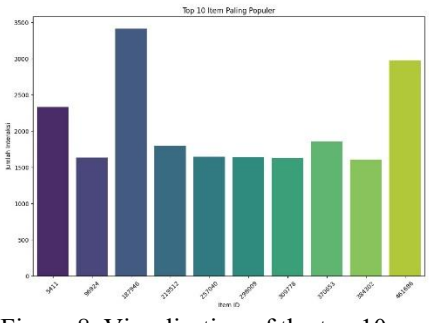

Figure 8. Visualisation of the top 10 most popular items

From the visualization of the "Top 10 Most Popular Items" it can be seen that there are several items that dominate user interaction. These items may have special characteristics, such as competitive prices, good

reviews, or ongoing promotions. Understanding more deeply about the factors that lead to the popularity of these items can help in designing more effective marketing strategies and product recommendations. In addition, this analysis can also help identify patterns of user behavior that can be used to increase sales.

4. CONCLUSION

In this study, we have conducted a series of analyses and visualizations of the Retailrocket Recommender System dataset to improve understanding of the e-commerce platform and associated item properties. Here are some key conclusions that can be drawn:

1. Item Category Distribution:
Based on the visualization of the distribution of item categories, we see that some categories have a more dominant number of items compared to other categories. This shows a significant difference in the distribution of items between these categories.

2. Distribution of User Interaction:
The distribution of user interactions shows that the majority of users perform a small number of interactions, while there are some very active users with a much larger number of interactions. This pattern is consistent with user behavior on e-commerce platforms where a small percentage of users tend to be more active compared to others.

3. Top 10 Most Popular Items:
Through the analysis of the top 10 most popular items, we can identify the items that users see or interact with most often. This information can be the basis for further analysis of user preferences and product marketing strategies.

4. Correlation Heatmap between Item Properties:
Analysis of the correlation heatmap between item properties reveals the existence of some relationship or correlation between item properties. For example, we can see that there is a positive correlation between properties A and B, while there is a negative correlation between properties C and D. This information can be used to understand the characteristics and relationships between items in the dataset.

Thus, this overall analysis and visualization provides valuable insights into item characteristics and user behavior in the context of e-commerce platforms. These results can be used as a basis for the development of more effective product recommendation strategies and a better understanding of user needs.

5. BIBLIOGRAPHY

[1] Ali Arifin. (2022). "Penerapan Sistem Algoritma Collaborative Filtering Untuk Rekomendasi Pemilihan Indekos Berdasarkan Rating." Teknologipintar.org, Vol. 2 (6).

[2]Ahmad Syaifuddin. (2023). "SISTEM REKOMENDASI PRODUK BERBAHASA INDONESIA PADA MARKETPLACE TOKOPEDIA MENGGUNAKAN METODE CONTENT BASE FILTERING." Jurnal Ilmiah Teknologi Informasi dan Sains Vol. 3 No.1

[3]Nurini Siregar, Samsudin. (2023). "Implementation of Collaborative Filtering Algorithms in Mobile Based Food Menu Ordering and Recommendation Systems" JURNAL MEDIA INFORMATIKA BUDIDARMA Vol. 7 No.3

[4] Hajaroh, Tati Suprapti, Riri Narasati (2024). "IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN ULASAN PRODUK MAKANAN DAN MINUMAN DI TOKOPEDIA" JURNAL MEDIA INFORMATIKA BUDIDARMA Vol. 8 No.1

[5]DITA AISHA (2022). "SISTEM REKOMENDASI TOKO ONLINE MENGGUNAKAN ALGORITMA COLLABORATIVE FILTERING DAN CONTENT BASED FILTERsING"

[6] Wang Juan, Lan Yue-xin, Wu Chun-ying (2019). "Survey of Recommendation Based on Collaborative Filtering" Journal of Physics: Conference Series.