# A novel Based-Approach Composed of Clustering Algorithm & Cosine Similarity for Products Recommendation

Mohammed Abdullah Al-Hagery

Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia
(e-mail: hajry@qu.edu.sa)

**Abstract— There are huge tons of transactions being accomplished online every day. This implies that e-commerce is facing the problem of data and information overloads. While customers are shopping via websites, they spend a lot of time to search for the required products based on their needs. This problem can easily be alleviated by having an accurate recommendation system based on a strong algorithm and confident measures in this regard. There are two main techniques for products recommendation; content-based filtering and collaborative filtering. If one of these two techniques implemented on the e-commerce system, a lot of limitations and weak points will appear. This paper aims at generating an optimal list of product, which, in turn, generates an accurate and reliable list of items. The new approach is composed of three components; clustering algorithm, user-based collaborative filtering, and the Cosine similarity measure. This approach implemented using a real dataset of past experienced users. The accuracy of the search results is a matter to users, it recommends the most appropriate products to users of the e-commerce website. This approach shows trustworthy results and achieved a high level of accuracy for recommending products to users.**

**Keywords—Recommendation Systems, Machine Learning, k-means clustering Algorithm, Collaborative Filtering, Similarity Measure, User Preferences.**

## I. INTRODUCTION

THE objective of transactions data analysis is to investigate and extracting hidden patterns from it by using diverse machine learning methods. Datasets are available in a large number of information industry repositories; these datasets are useless until it converted to helpful information then to knowledge. Therefore, it is fundamental to analyze this data and extract valuable knowledge from it. That permits clients to use data from various angles [1]. The recommendation system is an application of data mining concepts and machine learning algorithms. Nowadays, recommendation systems are important because it provides an easy way for users to find the products that they are seeking to buy.

Recommendation systems are used in many sites to recommend products to a set of customers [2], there are various applications of recommendation systems that are using different techniques [3]–[6]. Nowadays E-commerce websites are developing so quickly, for that it is a difficult action for online buyers to select a proper category. To deal with such a broad-ranging commercial problem, most electronic retailing sites merge the Internet services with buyer data to evolve a recommendation system, to predict their desire, they use buyers background and actions, then it helps E-commerce sites to make appropriate recommendations [7].

Often the search for a product on the sites is by writing keywords such as "red skirt, yellow t-shirt" or any other product or services. This search can be exhausting and annoying if the site does not have a useful Recommender System [8]. The recommendation systems assist users to discover items that meet their desires by recommending services or products to support users in making decisions in different domains, for example, what products a customer decides to buy, what film to look at, what paint to select, or what books are chosen to read. Making decisions on the websites is time-consuming, so by an effective recommendation system that will resolve the problem, save the time and effort when the user is looking for a particular product.

This paper is based on using a clustering and similarity approaches on the filtering process, which are content-based and collaborative filtering for getting an accurate list of recommended items. The importance of the proposed approach is to facilitate the search for a recommended products depends on various techniques works together, therefore the results will be more accurate.

The rest of the paper organized as follows: Section II explains the related works, section III demonstrates the methodology in details, and Section IV explains the results discussion, section V demonstrates the conclusions, finally, the future work in section VI.

## II. RELATED WORKS

The recommendation systems [9] are examples of web usage mining and web content mining, which depends on the study,

analysis of the web content, the user behavior, and preference of specific products over others during the navigation through the websites. Content recommendation system suggests a set of products to the customers based on the content of their buying history because it provides the content of the overview of the products, in which customer is generally attracted by many products. Finding out the quality of the item cannot be done via content-based filtering. To solve this issue, the cooperative filtering systems include "social filtering" or recommender system because they are based on the opinion of the other customers [10]. The strategy merging sentiment assessment with cooperative filtering as in [7], which used to enhance the recommendation results' accuracy of customers and attempts to overcome the problem of cold start and data sparsity.

Cooperative filtering systems applications provide customers with a good experience, but they still face some of the issues and challenges such as data sparsity. The data sparsity according to [11] is a very big problem and is one of the common issues because the accuracy of the recommendation with the cooperative filtering algorithm declines. There is a solution to overcome the data sparsity to enhance the accuracy and scalability of recommender systems, by using the clustering-based matrix factorization method. The idea is to find rating-patterns by mapping all customers/products to their corresponding customer/product clusters.

Building a recommendation system have a variety of methods such as content-based filtering, and traditional collaborative systems. For example, a Content-Based Filtering method has been successfully implemented using the techniques; Term Frequency (TF) and Inverse Document Frequency (IDF) [12], that show a comparison of data existing cosmetics in the database with the user input to increase their sales. Also, a type of clothing recommendation system has been presented [8], it based on the combination of visual features, visual attention, and textual attributes. The researchers conducted the tests on the system according to the recommendation and classification mechanisms. The outputs showed that their strategy reached good outputs when compared to the standard item k-nearest-neighbour method.

Recommendation systems can be used in different stores or sites like that in [1], where the researchers used a Recommender System, to recommend textbooks to help learners and people to find the best available books in the database that meet their preferences. Collaborative filtering and content-based filtering are used to find out the desired books based on ranking and content.

This method helped in solving the sparse data by applying three points including; Factorization, Clustering, and Approximation. According to [13], the collaborative filtering is implemented to obtain more accurate prediction based on user's preferences by combining of user's comments and scoring of the items. Likewise, some researchers, in their recommendation methods, employ the matrix-factorization procedures for learning customer preference from that matrix [14]–[16]. Recommendation systems are applied in different areas of our life, where one of the applications like [17], that provides a content-based mobile recipe that can recommend a customer

preferred a recipe using content-based filtering algorithm. Therefore, the recommendation systems helped to lighten the problem of overloaded information and allow the customers to access the relevant information and services according to [18]. There are a set of features, which may improve the quality and the impact of the collaborative filtering process identified in the suggested approach [19].

The recommendation system proposed by [20] minimizes the false-positive error that occurs frequently in the traditional system. The results proved that the accuracy achieved using an improved k-means algorithm were 82% to 85%, seems to be better than the old k-means algorithm. On the other hand, Zarzour et, al. proposed two various algorithms to improve the recommendation performance [21]. Beside that Chen et al. examined an attention function merged with the gated recurrent to increase the accuracy of the customer preference [22]. According to the research work achieved by [23], it obtained that the content recommender system limitation was in the early stage, which mainly related to insufficient data. The customer profile helped to obtain personalized content and recommend similar content products. So, this recommendation system was integrated with the content-based filtering method, which addressed the problem of the insufficient information and it became more efficient.

On the other hand, it found that trust is an important issue in building proper recommendation systems. Some researchers studied this issue in "TruCom company" that uses the social network for users in a specific domain to recommend the appropriate items for them [24]. Trust and secure recommendation system can solve many problems [25].

There are some recommendation systems make recommendations based on various approaches through different strategies [1], and other recommendation systems with developed by graph-based algorithm [26], or those systems designed by a concept level method [27]. Besides, in [28], many collaborative filtering strategies explained in personalized recommendation methods used for an E-commerce platform.

The collaborative filtering has been widely used in data mining and other fields such as that merge the user review text and rating [21], [29]. Also, a set of recommendation systems directed for recommending product items to users [30]–[32], although, it suffers some weaknesses.

Moreover, many types of recommendation systems emerged from the traditional types, such as the movies recommendation model, which proposed by [33]. This model can help customers by suggesting suitable movies based on their interests. It is a combination of a collaborative approach and content-based analysis. As well, similar types of these systems directed for films' purchasing purpose [34]–[36].

Besides, some researches focused on the establishment of a network of items depends on the history of users' item usage [37]. While some studies concentrated on the improvement of customer satisfaction [38] and others focused on customer rating for providing suitable recommendations [38]. And therefore, each one of the filtering strategies has its weakness and drawbacks.

On the other hand, to solve the information overload problem of products on the web and to make it helpful for all customers, it is imperative to establish an understanding of how content-based and collaborative filtering techniques work. The content-based filtering the suggested items to the buyer. This process carried out based on their experience with the system, which involves product ratings and previously purchased items. There are many recommendation systems developed to incorporate the technique TF and IDF must be accounted as they are essential in information retrieval systems. While TF is the definition of how frequent a word appears, IDF defines how frequent the document holding the word appears in the whole collection of documents. TF/IDF is a Vector Space Model. It is just one of the preferred models for developing the recommendations system.

The other models include probabilistic models such as Decision Trees and Naïve Bayes Classifier [39]. The collaborative algorithms include neighbourhood-based

developing recommender systems is giving suggestions that are matched with the customer's interests. Therefore, Singh et al [42], put a cornerstone of a roadmap for future research concerned with recommendation systems.

The proposed approach in this paper is introduced to enhance the recommendation systems quality and it is an extension of the work in [43]. Consequently, this paper is focusing on the development of an advanced type of recommendation systems that improve the accuracy of the final list of the recommended items or products based on the calculation of similarity and using the clustering technique as a machine learning algorithm.

## III. THE METHODOLOGY

The proposed approach takes into consideration the similarity between customers to improve the accuracy of the final recommendation and make more choices. A novel approach proposed consists of three techniques; clustering, user-based collaborative filtering, and measuring the similarity
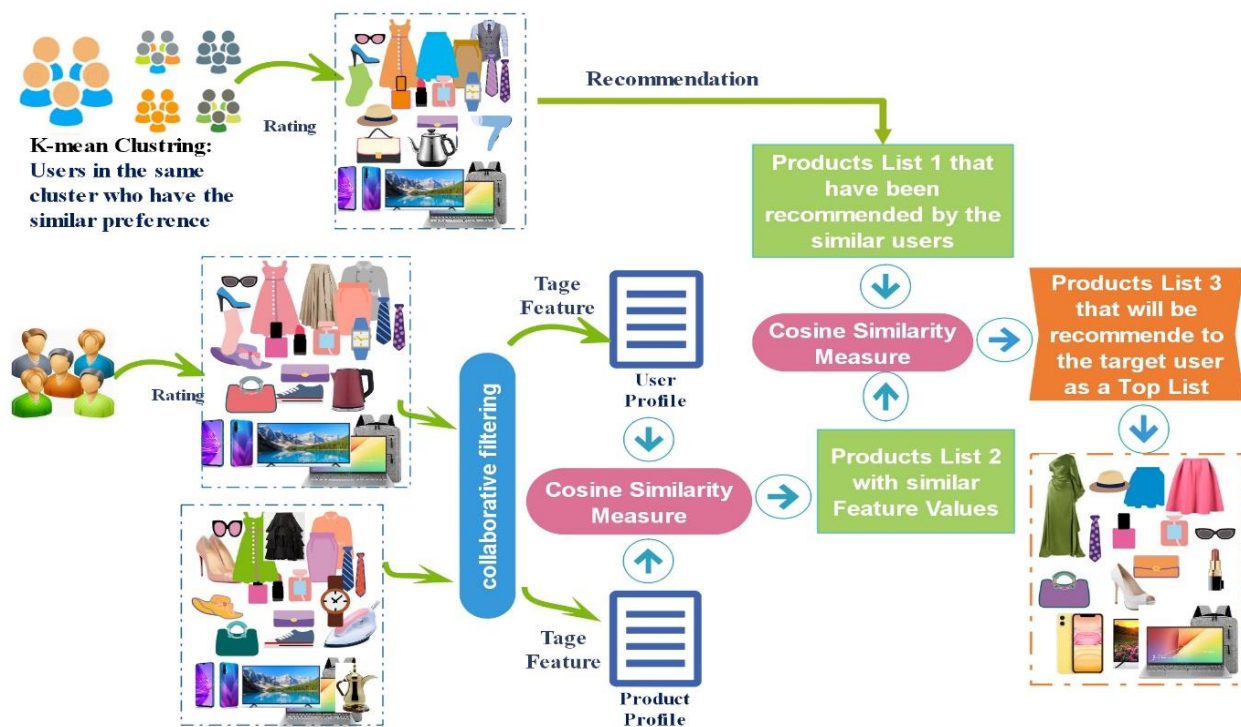


Fig. 1: The proposed approach general steps

algorithm and correspondence mean algorithm [40].

Currently, researchers are tried to improve the process, the quality, and the trusts of recommendations, for instance, Dong et al, introduced a systemic review of three classes of recommendation systems, that study leverage users' trust subject given by these recommender systems and they also discussed the ideas of enabling recommendation systems to give explanations, reasons and justifications to attract customers choose the recommended products [41], however, the process of providing explanations and justifications for choosing a product over another is difficult. The idea of

between items, using the cosine measure to obtain the final recommended products as shown in Figure 1.

This figure shows the whole processes of the approach. First, clustering users based on collaborative filtering. Second, measure the similarity between the products that will be recommended by a similar user. Third, select the top-recommended products that have similar features to the user preferences as in [44], but with high accuracy of selection.

The methodology implemented using a sample of previous experienced user's dataset gathered from the official repository of Kaggle. This methodology encompasses the following steps:

1. Clustering: by dividing customers based on their personal information to include similar users in the same cluster.
2. Collaborative filtering: by measuring the similarity between the users based on their preferences.
3. Final Recommended Items List: this step achieved by measuring the similarity between the recommended items where the similarity distance between the top-recommended products resulted from the original approach and recommended items of the proposed approach is calculated using cosine measure. The next section shows the implementation of these techniques.

### A. Clustering

Clustering is a grouping set of customers in similar groups, where the same cluster contains customers are more similar to each other than to those in other clusters. By this technique, the users divided based on two attributes (age and income) into four clusters using the K-mean algorithm. A sample of personal information of customers used to categorize similar customers together (Table 1). The sample contains various attributes, such as the age, income, occupation sex race, etc.

Table 1: A sample of Users' Personal Information

| Cust Id | Age | Income | Weight | Occupation | Sex | Race |
|---|---|---|---|---|---|---|
| 1 | 22 | 2,000 | 70 | Writer | F | w |
| 2 | 44 | 10,000 | 55 | Banker | M | b |
| 3 | 34 | 9,000 | 44 | Nurse | F | b |
| 4 | 67 | 1,500 | 67 | Singer | F | a |
| 5 | 65 | 1,000 | 63 | teacher | M | a |
| 6 | 32 | 5,000 | 89 | biologist | F | i |
| 7 | 44 | 8,000 | 50 | chemical | M | w |
| 8 | 20 | 2,500 | 67 | Teacher | F | b |
| 9 | 30 | 4,500 | 76 | Nurse | F | a |
| 10 | 35 | 5,000 | 87 | Doctor | M | i |
| 11 | 22 | 3,000 | 65 | Waiter | M | w |
| 12 | 19 | 1,500 | 67 | Singer | M | b |
| 13 | 32 | 4,200 | 97 | Nurse | F | a |
| 14 | 27 | 6,000 | 78 | Teacher | F | i |
| 15 | 50 | 3,000 | 87 | Banker | M | w |
| 16 | 45 | 9,000 | 76 | Writer | M | b |
| 17 | 55 | 10,000 | 98 | Teacher | F | a |
| 18 | 39 | 8,500 | 78 | Doctor | M | w |
| 19 | 23 | 2,000 | 76 | Writer | F | b |
| 20 | 52 | 9,500 | 87 | Chemical | F | a |
| 21 | 15 | 1,000 | 62 | Singer | F | i |
| 22 | 28 | 6,000 | 83 | Waiter | M | w |
| 23 | 26 | 5,500 | 92 | Banker | F | b |
| 24 | 33 | 7,000 | 100 | Writer | F | a |
| 25 | 38 | 7,500 | 69 | Nurse | F | i |
| 26 | 47 | 9,000 | 56 | Doctor | M | w |
| 27 | 16 | 800 | 76 | Biologist | F | b |
| 28 | 46 | 9,000 | 87 | Manger | M | a |
| 29 | 36 | 4,000 | 84 | Singer | M | i |
| 30 | 41 | 2,000 | 57 | Doctor | F | w |

Clustering was applied using the SPSS software to find clustering results based on two steps; First, choose the number of clusters K, where K=4. Second, run the clustering task based on two attributes; age and income (Table 2). The clustering results are demonstrated (Tables 2 to 7). The initial cluster centres are evaluated by applying the first estimate of the variable of the means to find the K centres for each cluster. Initial cluster centres are used for the first round of clustering then will be updated (Table 2).

Table 2: Initial Cluster Centers

| | Clusters | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| age | 38 | 55 | 36 | 16 |
| income | 7500 | 10000 | 4000 | 800 |

The final centre of each cluster is generated as a mean of each variable in the final content of each cluster. The final cluster centres denote the features of the typical case for each cluster (Table 3).

Table 3: Final Cluster Centers

| | Clusters | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Age | 35 | 46 | 31 | 34 |
| Income | 7167 | 9357 | 4078 | 1475 |

The distance between the final cluster centres carried out by using the Euclidean distances and the larger distances between clusters centre result of greater dissimilarities. Where the most different clusters are cluster 2 and cluster 4. Also, when the clusters are compared to itself, the result of the difference is zero, as the distance between cluster 1 and 1 (Table 4).

Table 3: Distance between final cluster centers

| Cluster | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0 | 2190.505 | 3088.891 | 5691.667 |
| 2 | 2190.505 | 0 | 5279.386 | 7882.153 |
| 3 | 3088.891 | 5279.386 | 0 | 2602.779 |
| 4 | 5691.667 | 7882.153 | 2602.779 | 0 |

The number of cases allocated to each cluster is presented (Table 5), the large number of customers included in the third group have an average income.

Table 5: The number of Cases per Cluster

| Cluster | 1 | 6.000 |
|---|---|---|
| | 2 | 7.000 |
| | 3 | 9.000 |
| | 4 | 8.000 |
| Valid | | 30.000 |
| Missing | | 0.000 |

Customers categorized into four groups based on their information records (Table 6). It includes a clustered column that is referred to as the cluster category for each customer, distances column that shows the distance between the initial

cluster centroids and the object in the same cluster, i.e., customer 17 is very close to the initial cluster centre of cluster 2, where the age is 55 and the income is; 10000$.

Table 4: Categorization of Customers

| Customer_ID/ Case Number | Cluster | Distance |
|---|---|---|
| 7 | 1 | 500.036 |
| 14 | | 1500.040 |
| 18 | | 1000 |
| 22 | | 1500.033 |
| 24 | | 500.025 |
| 25 | | 0.0 |
| 2 | 2 | 11 |
| 3 | | 1000.220 |
| 16 | | 1000.050 |
| 17 | | 0.0 |
| 20 | | 500.009 |
| 26 | | 1000.032 |
| 28 | | 1000.040 |
| 6 | 3 | 1000.008 |
| 8 | | 1500.085 |
| 9 | | 500.036 |
| 10 | | 1000 |
| 11 | | 1000.098 |
| 13 | | 200.040 |
| 15 | | 1000.098 |
| 23 | | 1500.033 |
| 29 | | 0.0 |
| 1 | 4 | 1200.015 |
| 4 | | 701.855 |
| 5 | | 205.915 |
| 12 | | 700.006 |
| 19 | | 1200.020 |
| 21 | | 200.002 |
| 27 | | 0.000 |
| 30 | | 1200.260 |

## B. Collaborative Filtering

The user-based collaborative filtering used by applying similarity computation to find a set of other users whose ratings are similar to target user using centred cosine that expressed in equation (1).

$$Sim(A,B) = corr(A,B) = \frac{\sum_{u \in U}(R_{u,A}-R_A)(R_{u,B}-R_B)}{\sqrt{\sum_{u \in U}(R_{u,A}-R_A)^2}\sqrt{\sum_{u \in U}(R_{u,B}-R_B)^2}} \quad (1)$$

Where:
- $R_{u,A}$ represents the rating of user u for item A.
- $R_A$ is the average rating of item A.
- $R_{u,B}$ represents the rating value of user u of item B.
- $R_B$ denotes average rating of item B.

A sample of user rating is given using the product-rating matrix (Table 7). Each row in this table/matrix represents a user who has rating different products, where rating start from zero (for low rating) to five (for high rating). The empty fields refer

to the user does not have been making a rating for this product. The implementation carried out in two steps. Normalizing the rating data then calculating similarity distance between users based on their rating.

Table 5: Sample of Customers Products Rating

| Customer_id | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
|---|---|---|---|---|---|
| 1 | - | 2 | 3 | 4 | 1 |
| 2 | - | - | 3 | - | 4 |
| 3 | 3 | - | 2 | - | - |
| 4 | 1 | 2 | - | 1 | 3 |
| 5 | - | 2 | - | - | 2 |
| 6 | 5 | - | 3 | 4 | 1 |
| 7 | - | 1 | 3 | - | - |
| 8 | 5 | 2 | - | 5 | - |
| 9 | 1 | - | 2 | 2 | 5 |
| 10 | - | 4 | - | - | 5 |
| 11 | 2 | - | 4 | 2 | 1 |
| 12 | - | - | 4 | - | 1 |
| 13 | 1 | - | 2 | 1 | 5 |
| 14 | 4 | - | 1 | 1 | - |
| 15 | 1 | - | 3 | 2 | 1 |

The first step output is illustrated (Table 8). The user-based collaborative filtering is applied, where rating data is normalized by subtracting row means to get the modified rating matrix. The main purpose of the normalization is to avoid calculating the products that have not been rating as a low rating.

Table 6: Normalization of the Customer Rating

| Customer_id | Item1 | Item2 | Item3 | Item4 | Item5 |
|---|---|---|---|---|---|
| 1 | - | -0.5 | 0.5 | 1.5 | -1.5 |
| 2 | - | - | -0.5 | - | 0.5 |
| 3 | 0.5 | - | -0.5 | - | - |
| 4 | -0.75 | 0.25 | - | -0.75 | 1.25 |
| 5 | - | 1 | - | - | 1 |
| 6 | 1.75 | - | -0.25 | 0.75 | -2.25 |
| 7 | - | -1 | 1 | - | - |
| 8 | 1 | -2 | - | 1 | - |
| 9 | -1.5 | - | -0.5 | -0.5 | 2.5 |
| 10 | - | -0.5 | - | - | 0.5 |
| 11 | -1 | - | 1 | -1 | -2 |
| 12 | - | - | 1.5 | - | -1.5 |
| 13 | -1.25 | - | -0.25 | -1.25 | 2.75 |
| 14 | 2 | - | -1 | -1 | - |
| 15 | -0.75 | - | 1.25 | 0.25 | -0.75 |

## C. Final Recommended items List

When the similarity degree is increased then the distance will be reduced [45]. As the second step, the similarity between users based on the Pearson Correlation Coefficient [46], calculated using equation (2).

$$Sim(a,b) = \frac{\sum_{p \in P}(r_{a,p}-\overline{r_a})(r_{b,p}-\overline{r_b})}{\sqrt{\sum_{p \in P}(r_{a,b}-\overline{r_a})^2}\sqrt{\sum_{p \in P}(r_{b,p}-\overline{r_b})^2}} \quad (2)$$

Where:

- a, b refers to the users.
- p={$p_1, \ldots, p_m$}, denotes the sets of products.
- $r_{a,p}$: denotes the rating of the user a for product p.
- $\overline{r_a}$: refers to the average rating of the user a.

The similarity distance between fifteen users (Table 9) demonstrated based on their product rating (Table 7), where great distances between user testes correspond to higher similarities.

find similar items (based on the product description) that will be also recommended to the target user. A sample of a product profile matrix (Table 10), where there are six products, three given by the first approach, namely item A, item B, and item C. Three given by the proposed approach, namely item 1, item 2, and item 3 and the combined features of these products are 'short', 'occasion', 'skinny', and bright. The values of Item A, Item B, Item C, Item 1, Item 2, and Item 3 were normalized by

Table 7: The Similarity between Users Based on Products Rating

| User 1 | User 2 | User 3 | User 4 | User 5 | User 6 | User 7 |
|---|---|---|---|---|---|---|
| Sim(1,2)=-0.632 | Sim(2,3)=0.5 | Sim(3,4)=-0.319 | Sim(4,5)=0.825 | Sim(5,6)=-0.694 | Sim(6,7)=-0.059 | Sim(7,8)=0.577 |
| Sim(1,3)= -0.158 | Sim(2,4)=0.533 | Sim(3,5)=0 | Sim(4,6)=-0.955 | Sim(5,7)=-0.645 | Sim(6,8)=0.345 | Sim(7,9)=-0.117 |
| Sim(1,4)= -0.842 | Sim(2,5)=0.645 | Sim(3,6)=0.478 | Sim(4,7)=-0.1066 | Sim(5,8)=0.745 | Sim(6,9)=-0.957 | Sim(7,10)=0.5 |
| Sim(1,5)=-0.816 | Sim(2,6)=-0.48 | Sim(3,7)=-0.5 | Sim(4,8)=-492 | Sim(5,9)=0.760 | Sim(6,10)=-0.537 | Sim(7,11)=0.31 |
| Sim(1,6)=0.661 | Sim(2,7)=-0.5 | Sim(3,8)=0.288 | Sim(4,9)=0.929 | Sim(5,10)=0 | Sim(6,11)=0.259 | Sim(7,12)=0.5 |
| Sim(1,7)=0.316 | Sim(2.8)=0 | Sim(3,9)=-0.235 | Sim(4,10)=0.426 | Sim(5,11)=-0.32 | Sim(6,12)=0.478 | Sim(7,13)= - 0.05 |
| Sim(1,8)=0.457 | Sim(2,9)=0.707 | Sim(3,10)=0 | Sim(4,11)=-0.264 | Sim(5,12)=-0.645 | Sim(6,13)=0.462 | Sim(7,14)=0.29 |
| Sim(1,9)=-0.708 | Sim(2,10)=0.5 | Sim(3,11)=-0.62 | Sim(4,12)= -0.533 | Sim(5,13)=0.765 | Sim(6,14)=-0.432 | Sim(7,15)=0.53 |
| Sim(1,10)=-0.316 | Sim(2,11)= -0.93 | Sim(3,12)=-0.5 | Sim(4,13)=0.977 | Sim(5,14)=0 | Sim(6,15)=0.168 | - |
| Sim(1,11)=0.392 | Sim(2,12)=-1 | Sim(3,13)=-0.22 | Sim(4,14)=-0.184 | Sim(5,15)= -0.41 | - | - |
| Sim(1,12)=0.632 | Sim(2,13)=0.65 | Sim(3,14)=0.866 | Sim(4,15)=-0.204 | - | - | - |
| Sim(1,13)=-0.835 | Sim(2,14)=0.29 | Sim(3,15)=-0.852 | - | - | - | - |
| Sim(1,14)=-0.365 | Sim(2,15)= -.85 | - | - | - | - | - |
| Sim(1,15)=0.573 | - | - | - | - | - | - |
| **User 8** | **User 9** | **User 10** | **User 11** | **User 12** | **User 13** | **User 14** |
| Sim(8,9)=-0.27 | Sim(9,10)=-0.3 | Sim(10,11)=-0.62 | Sim(11,12)=0.93 | Sim(12,13)= -0.65 | Sim(13,14)=-0.13 | Sim(14,15)=-0.74 |
| Sim(8,10)=0.58 | Sim(9,11)=-0.51 | Sim(10,12)= -0.5 | Sim(11,13)=-0.43 | Sim(12,14)= -0.29 | Sim(13,15)=-0.32 | - |
| Sim(8,11)=-0.36 | Sim(9,12)=-0.58 | Sim(10,13)=0.593 | Sim(11,14)=-0.36 | Sim(12,15)=0.85 | - | - |
| Sim(8,12)=0 | Sim(9,13)=0.965 | Sim(10,14)=0 | Sim(11,15)=0.86 | - | - | - |
| Sim(8,13)=-0.31 | Sim(9,14)=-0.27 | Sim(10,15)=-0.32 | - | - | - | - |
| Sim(8,14)=0.17 | Sim(9,15)=-0.3 | - | - | - | - | - |
| Sim(8,15)=-0.12 | - | - | - | - | - | - |

Let consider user 9 and user 13 are target users, it notices that the two users have more similar taste and they are in the same cluster, so the products that preferred by user 13 will be preferred and recommended too by user 9.

dividing the term occurrence 1 or 0 (Table 10), by sorting the number of attributes in the item description, the results are given (Table 11).

Table 8: The Items Resulted from Both Approaches

| Item Name | Bright | Skinny | Occasion | Short |
|---|---|---|---|---|
| Item A | 1 | - | 1 | - |
| Item B | 1 | 1 | 1 | - |
| Item C | 1 | - | - | 1 |
| Item 1 | 1 | 1 | 1 | - |
| Item 2 | 1 | - | - | 1 |
| Item 3 | 1 | - | 1 | 1 |

Measuring similarity between items that are given by two approaches based on their description, by taking the top-recommended items that are a result of the first approach and

Table 9: The values of the Normalized Items

| Item Name | Bright | Skinny | Occasion | Short |
|---|---|---|---|---|
| Item A | 0.7071 | - | 0.7071 | - |
| Item B | 0.5773 | 0.5773 | 0.5773 | - |
| Item C | 0.7071 | - | - | 0.7071 |
| Item 1 | 0.5773 | 0.5773 | 0.5773 | - |
| Item 2 | 0.7071 | - | - | 0.7071 |
| Item 3 | 0.5773 | - | 0.5773 | 0.5773 |

The similarity between the items measured using centered cosine by equation (1). In the original approach, the results represented by product B, A, and C respectively. So, the product B will be put first then finding the similar products that

given by the proposed approach, after that, item A is taken after the calculation of their similarity and so on so forth with the other cases. For item B, it found that item 1 is the most similar item to B by calculating the similarities: Sim(B,1) = 1. Therefore, item 1 will be the next recommended item after item B. The same process can be applied to other items in the dataset.

## IV. RESULTS DISCUSSION

An increased accuracy of search results that leads to user satisfaction depends on the assumption that users are a neighborhood have similar tests and that user preferences never change. With a hybrid recommender system in place, expected results for products recommendations will be optimal due to increased accuracy, resulted from the multiple stages of filtering depends on the user preference and rating, in addition to that, the use the similarity measure and the collaborative filtering.

Consequently, through the research results, it found that the proposed approach is working perfectly and producing accurate results, some examples will be discussed here clearly, for instance, let take in consideration user 5 is a target user, first, the recommendation system will search for all products that are similar to the products that user 5 has been rating. The second is the system will search for the similar users to user 5 by calculating the similarity, it will found that the user 4 is the most similar user to user 5 where the similarity rate between user 4 and user 5 is high to some extent (0.825%) and both are in the same cluster (cluster 4).

In the same time, it has been found that user 9 and user 11 in the same cluster (cluster 3), so the products that have been rating by user 9 are recommended to user 13. The output items will be listed as follows:

Product 2 (Item 2), Product 1 and product 3. Third, the recommendation system will calculate the similarities between all the output results. Where the most similar product to the product A (let take product 1 for example) is listed after product A.

Likewise, another example can be given a good similarity rate, as follows: let take in consideration user 13 is a target user. First, the recommendation system will search for the products that are similar to the products that user 13 has been rating. The output items will be listed as follows: product A, product B, and product C. The second is the system will search for the similar users to user 13 by similarity computation, it will be found that the user 9 is the most similar user to user 13 where the similarity value between user 9 and user 13 is very high (0.965%).

## V. CONCLUSIONS

The research finding is significant and necessary for a large segment of customers in society, especially those are preferring to buy their needs through E-commerce. People in general, are attracted to the sites that have a powerful recommendation system because it makes it easier for them in searching. The proposed approach involves the clustering algorithm and user-based collaborative filtering to recommend the products that are favourable by a user to other users who has a similar preference of the same or similar products.

Then measuring the similarity between the products to recommend the most similar one for them. The paper also addresses some of the critical concerns in product recommendation such as insufficient data for recommendation and novelty. Factorization, clustering, and approximation as perceived as the constructs of a better recommender system capable of addressing the issue of insufficient data for the recommendation. The issue of novelty, which mainly arises from collaborative filtering techniques, can be overcome by measuring the similarity between recommended items and using clustering hence maximizing the significance of the proposed system as a new hybrid method has the ability for achieving the required goals. Furthermore, the proposed approach shows trustworthy results and is achieving a high level of accuracy for recommending products to users according to the multi-stages selection.

## VI. FUTURE WORK

The future works can focus on improving the product recommendation process by giving justifications for answering why a list of items is recommended. The proposed approach can be enhanced by making modifications to its current structure by using other algorithms such as the deep learning algorithms in the side of users and products clustering. Besides, the filtration process of products can be improved by suggesting new filters, as well the similarity calculations can be measured by using other statistical measures such as the Manhattan distance, Minkowski distance, and Jaccard similarity.

## REFERENCES

[1] P. Mathew, B. Kuriakose, and V. Hegde, "Book Recommendation System through content based and collaborative filtering method," in *Proceedings of 2016 International Conference on Data Mining and Advanced Computing, SAPIENCE 2016*, 2016, pp. 47–52, doi: 10.1109/SAPIENCE.2016.7684166.

[2] A. S. Tewari and K. Priyanka, "Book recommendation system based on collaborative filtering and association rule mining for college students," in *Proceedings of 2014 International Conference on Contemporary Computing and Informatics, IC3I 2014*, 2014, pp. 135–138, doi: 10.1109/IC3I.2014.7019651.

[3] M. A. Hameed, O. Al Jadaan, R. Al-khaimah, U. A. Emirates, and S. Ramachandram, "Collaborative Filtering Based Recommendation System: A survey," *Int. J. Comput. Sci. Eng. Collab.*, vol. 4, no. 5, pp. 859–876, 2012.

[4] G. Geetha, M. Safa, C. Fancy, and D. Saranya, "A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System," *J. Phys. Conf. Ser.*, vol. 1000, no. 1, 2018, doi: 10.1088/1742-6596/1000/1/012101.

[5] M. D. Ekstrand, "Collaborative Filtering Recommender Systems," *Found. Trends® Human–Computer Interact.*, vol. 4, no. 2, pp. 81–173, 2011, doi: 10.1561/1100000009.

[6] A. E. Patil, S. Patil, ; Karanjit Singh, P. Saraiya, and A. Sheregar, "Online Book Recommendation System Using Association Rule Mining And Collaborative Filtering," *Int. J. Comput. Sci. Mob. Comput.*, vol. 8, no. 4, pp. 83–87, 2019.

[7] N. Jing, T. Jiang, J. Du, and V. Sugumaran, "Personalized recommendation based on customer preference mining and sentiment assessment from a Chinese e-commerce website," *Electron. Commer. Res.*, vol. 18, no. 1, pp. 159–179, 2018, doi: 10.1007/s10660-017-9275-6.

[8] E. Viriato De Melo, E. A. Nogueira, and D. Guliato, "Content-based filtering enhanced by human visual attention applied to clothing recommendation," in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2016, vol. 2016-Janua, pp. 644–651, doi: 10.1109/ICTAI.2015.98.

[9] R. Baraglia and F. Silvestri, "Dynamic personalization of Web sites without user intervention," *Commun. ACM*, vol. 50, no. 2, pp. 63–67, 2007, doi: 10.1145/1216016.1216022.

[10] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating 'word of mouth,'" in *Conference on Human Factors in Computing Systems - Proceedings*, 1995, vol. 1, pp. 210–217.

[11] Z. Shajarian, S. A. Seyedi, and P. Moradi, "A clustering-based matrix factorization method to improve the accuracy of recommendation systems," in *Electrical Engineering (ICEE), 2017 Iranian Conference on*, 2017, pp. 2241–2246, doi: 10.1109/iraniancee.2017.7985435.

[12] J. C. Patty, E. T. Kirana, M. Sandra, and D. Khrismayanti, "Recommendations System for Purchase of Cosmetics Using Content- Based Filtering," vol. 10, no. 1, pp. 1–5, 2018.

[13] Y. Zhu, W. Song, L. Liu, X. Zhao, and C. Du, "Collaborative filtering recommender algorithm based on comments and score," in *Proceedings - 2017 10th International Symposium on Computational Intelligence and Design, ISCID 2017*, 2017, vol. 1 2018-Jan, pp. 304–307, doi: 10.1109/ISCID.2017.158.

[14] G. Hinton, "Restricted Boltzmann Machines for Collaborative Filtering Ruslan," *J. Comput. Inf. Technol.*, vol. 6, no. 3, pp. 245–254, 1998.

[15] M. Quadrana, A. Karatzoglou, B. Hidasi, and P. Cremonesi, "Personalizing Session-based Recommendations with Hierarchical Recurrent Neural Networks," 2017, doi: 10.1145/3109859.3109896.

[16] B. Joeran and G. Bela, S. Langer and C. Breitinger, "Research Paper Recommender Systems : A Literature Survey Table of Content," *Int. J. Digit. Libr.*, vol. 17, November 2016, pp. 305–338, 2016, doi: 10.1007/s00799-015-0156-0.

[17] T. L. Cheng, U. K. Yusof, and M. N. A. Khalid, "Content-based filtering algorithm for mobile recipe application," in *2014 8th Malaysian Software Engineering Conference, MySEC 2014*, 2014, pp. 183–188, doi: 10.1109/MySec.2014.6986011.

[18] S. Ouaftouh, A. Zellou, and A. Idri, "CARMS: Clustering based approach for recommendation in mediation systems," in *Proceedings - 2017 International Conference on Innovative and Creative Information Technology: Computational Intelligence and IoT, ICITech 2017*, 2018, vol. 2018-Janua, pp. 1–6, doi: 10.1109/INNOCIT.2017.8319152.

[19] L. M. T. Schouten, R. P. T. M. Grol, and M. E. J. L. Hulscher, "Factors influencing success in quality-improvement collaboratives: Development and psychometric testing of an instrument," *Implement. Sci.*, vol. 5, no. 1, p. 84, 2010, doi: 10.1186/1748-5908-5-84.

[20] A. Saklecha and J. Raikwal, "Dynamic Recommendation System Using Enhanced K-means Clustering Algorithm for E-commerce," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 5, pp. 159–163, 2017.

[21] H. K. Lam, S. S. H. Ling, and H. T. Nguyen, *Computational Intelligence and Its Applications*, vol. 1. Springer International Publishing, 2012.

[22] J. Chen and A. Abdul, "A Session-Based Customer Preference Learning Method by Using the Gated Recurrent Units with Attention Function," *IEEE Access*, vol. 7, pp. 17750–17759, 2019, doi: 10.1109/ACCESS.2019.2895647.

[23] W. Paireekreng, "Mobile content recommendation system for re-visiting user using content-based filtering and client-side user profile," in *Proceedings - International Conference on Machine Learning and Cybernetics*, 2013, vol. 4, pp. 1655–1660, doi: 10.1109/ICMLC.2013.6890864.

[24] H. Liu, F. Xia, Z. Chen, N. Y. Asabere, J. Ma, and R. Huang, "TruCom: Exploiting Domain-Specific Trust Networks for Multicategory Item Recommendation," *IEEE Syst. J.*, vol. 11, no. 1, pp. 295–304, 2017, doi: 10.1109/JSYST.2015.2427193.

[25] B. Ramesh and R. Reeba, "Secure recommendation system for E-commerce website," in *Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2017*, 2017, pp. 1–5, doi: 10.1109/ICCPCT.2017.8074240.

[26] S. Shaikh, S. Rathi, and P. Janrao, "Recommendation system in E-Commerce Websites: A graph based approached," in *Proceedings - 7th IEEE International Advanced Computing Conference, IACC 2017*, 2017, pp. 931–934, doi: 10.1109/IACC.2017.0189.

[27] N. S. B. C. Lah, A. R. B. C. Hussin, and H. M. Dahlan, "A concept-level approach in analyzing review readership for E-Commerce persuasive recommendation," in *International Conference on Research and Innovation in Information Systems, ICRIIS*, 2017, pp. 1–5, doi: 10.1109/ICRIIS.2017.8002443.

[28] L. Yan, "Personalized recommendation method for E-commerce platform based on data mining technology," in *Proceedings - 2017 International Conference on Smart Grid and Electrical Automation, ICSGEA 2017*, 2017, vol. 2017-Janua, pp. 514–517, doi: 10.1109/ICSGEA.2017.62.

[29] X. Wang and C. Wang, "Recommendation system of e-commerce based on improved collaborative filtering algorithm," *Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS*, vol. 2017-Novem, pp. 332–335, 2018, doi: 10.1109/ICSESS.2017.8342926.

[30] J. Xia, "E-Commerce product recommendation method based on collaborative filtering technology," in *Proceedings - 2016 International Conference on Smart*

*Grid and Electrical Automation, ICSGEA 2016*, 2016, pp. 90–93, doi: 10.1109/ICSGEA.2016.81.

[31] Y. Liu and J. Du, "Parameter Optimization of the SVM for Big Data," *Proc. - 2015 8th Int. Symp. Comput. Intell. Des. Isc. 2015*, vol. 2, no. 1, pp. 341–344, 2016, doi: 10.1109/ISCID.2015.185.

[32] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th International Conference on World Wide Web, WWW 2001*, 2001, pp. 285–295, doi: 10.1145/371920.372071.

[33] F. Furtado, "Career Recommendation System Using Machine Learning Algorithm," vol. 9, no. 1, pp. 84–98, 2020, doi: 10.22105/riej.2020.226178.1128.

[34] C. C. Aggarwal and C. X. Zhai, "A survey of text classification algorithms," in *Mining Text Data*, vol. 9781461432, C. C. Aggarwal and C. Zhai, Eds. Boston, MA: Springer US, 2012, pp. 163–222.

[35] T. Zhou, L. Chen, and J. Shen, "Movie Recommendation System Employing the User-Based CF in Cloud Computing," in *International Conference on Computational Science and Engineering*, 2017, vol. 2, pp. 46–50, doi: 10.1109/CSE-EUC.2017.194.

[36] A. Pal, P. Parhi, and M. Aggarwal, "An improved content based collaborative filtering algorithm for movie recommendations," in *2017 10th International Conference on Contemporary Computing, IC3 2017*, 2018, vol. 2018-Janua, pp. 1–3, doi: 10.1109/IC3.2017.8284357.

[37] T. Ha and S. Lee, "Item-network-based collaborative filtering: A personalized recommendation method based on a user's item network," *Inf. Process. Manag.*, vol. 53, no. 5, pp. 1171–1184, 2017, doi: 10.1016/j.ipm.2017.05.003.

[38] Y. Jiang, J. Shang, and Y. Liu, "Maximizing customer satisfaction through an online recommendation system: A novel associative classification model," *Decis. Support Syst.*, vol. 48, no. 3, pp. 470–479, 2010, doi: 10.1016/j.dss.2009.06.006.

[39] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egypt. Informatics J.*, vol. 16, no. 3, pp. 261–273, 2015.

[40] C. L. Hee, J. L. Seok, and J. C. Young, "A study on the improved collaborative filtering algorithm for recommender system," in *Proceedings - SERA 2007: Fifth ACIS International Conference on Software Engineering Research, Management, and Applications*, 2007, pp. 297–304, doi: 10.1109/SERA.2007.33.

[41] M. Dong, F. Yuan, L. Yao, X. Wang, X. Xu, and L. Zhu, "Trust in Recommender Systems: A Deep Learning Perspective," pp. 1–34, 2020, doi: arXiv preprint arXiv:2004.03774.

[42] P. K. Singh, P. Kanti, D. Pramanik, A. K. Dey, and P. Choudhury, "Recommender Systems : An Overview , Research Trends and Future Directions," *Int. J. Bus. Syst. Res.*, January, 2020.

[43] T. Badriyah, E. T. Wijayanto, I. Syarif, and P. Kristalina, "A hybrid recommendation system for E-commerce based on product description and user profile," in *7th International Conference on Innovative Computing Technology, INTECH 2017*, 2017, Intech, pp. 95–100, doi: 10.1109/INTECH.2017.8102435.

[44] W. Zhou and W. Han, "Personalized recommendation via user preference matching," *Inf. Process. Manag.*, vol. 56, no. 3, pp. 955–968, 2019, doi: 10.1016/j.ipm.2019.02.002.

[45] S. Bandyopadhyay and S. Saha, *Unsupervised classification: Similarity measures, classical and metaheuristic approaches, and applications*, Kindle Edi. Springer, 2013.

[46] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, "Applying Collaborative Filtering to Usenet News," *Commun. ACM*, vol. 40, no. 3, pp. 77–87, 1997, doi: 10.1145/245108.245126.

## Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)