Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 10, No. 1, 2022, 103-116.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

# A Clustering-Classification Recommender System based on Firefly Algorithm

Hamidreza Koosha[1*], Zahra Ghorbani[2] and Roshanak Nikfetrat[2]

*1. Department of Industrial Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.*
*2. Department of Industrial Engineering, Sadjad University of Technology, Mashhad, Iran.*

**Article Info**

**Article History:**
*Received 30 April 2021*
*Revised 17 July 2021*
*Accepted 13 December 2021*

*DOI:10.22044/JADM.2021.10782.2216*

**Keywords:**
*Recommender System, Firefly Algorithm, K-Means, K-Nearest Neighbor, Naïve Bayesian, Clustering, Classification.*

**Corresponding author: koosha@um.ac.ir (H. Koosha).*

**Abstract**

In the last decade, online shopping has played a vital role in the customers' approach to purchase different products, providing convenience to the shops and many benefits for the economy. E-commerce is widely used for digital media products such as movies, images, and software. Thus, the recommendation systems are of great importance, especially in the today's hectic world, which searches for the content that would be interesting to an individual. This research proposes a new two-step recommender system based on the demographic data and user ratings on the public MovieLens datasets. In the first step, clustering on the training dataset is performed based on the demographic data, grouping customers in homogeneous clusters. The clustering includes a hybrid Firefly Algorithm (FA) and K-means approach. Due to the FA's ability to avoid trapping into the local optima, which resolves K-means' main pitfall, the combination of these two techniques leads to a much better performance. In the next step, for each cluster, two recommender systems are proposed based on K-Nearest Neighbor (KNN) and Naïve Bayesian Classification. The results obtained are evaluated based on many internal and external measures like the Davies-Bouldin index, precision, accuracy, recall, and F-measure. The results obtained show the effectiveness of the K-means/FA/KNN compared with the other extant models.

## 1. Introduction

With the rapid growth of the Internet, the amount of data transactions that happen by the ever-increasing number of users has significantly increased. However, the users often face the information overload problem, negatively affecting their productivity and the decision-making process. The researchers have suggested Recommendation Systems (RS) dealing with these problems. A recommendation system collects data about the user's preferences of different items, either implicitly or explicitly [1]. The recommender system models based on the types of input data are mainly classified into three categories: Collaborative Filtering (CF), Content-Based Filtering (CBF), and hybrid [2]. Also, the recommender systems have been deployed in various areas such as movies, shopping, tourism, and TV [3-6]. These days, most organizations implement the recommendation systems in order to fulfil the customer requirements, ranging from LinkedIn and Amazon to Netflix. For example, Netflix considers the types of shows that a customer watches and provides similar recommendations [7]. It has been shown that a useful application of RSs leads to the retention of the current users, and may help to acquire new ones [8, 9]. By increasing the efficiency and accuracy of the recommender systems, businesses can be expected to boost their revenues through customer acquisition and retention.

It is usually difficult for the users to find the appropriate movies aligned with their tastes, especially with the massive number of movies available worldwide. Different users like different movies or actors. Moreover, the business's end goal is to increase sales, revenues, user engagement or other metrics. Any inappropriate recommendation in this industry may affect the customers' loyalty, and may lead to customer churn [3]. Thus, it is essential to find a method of filtering irrelevant movies and find a set of relevant movies. A movie recommendation system analyzes the factors such as a review, cast, plot, crew, genre, and popularity. It helps the users to quickly search over the web to make decisions for the items related to their choice [1].

In the prior literature, numerous research works have studied the RS and sophisticated solutions for RS [3, 6, 11, 12]; yet there is still a gap between the models' expectations and performance. Many users' computational costs are also a critical issue [13]. Thus, there is a need to develop more effective RSs.

In this research work, a novel RS model is presented by considering the criteria such as accuracy, precision, and recall. In the proposed recommender system, two different data mining methods are used in two distinct stages in order to provide recommendations to the new user (target customer). This article has two main contributions: (1) the heuristic algorithm (FA-K-Means algorithm) clusters the data. This algorithm improves the common and practical K-Means algorithm to escape local optimal solutions; (2) two different methods are used in order to evaluate and compare the KNN and Naïve Bayesian methods. This research work develops and tests the proposed model based on a famous movie dataset. We selected a movie database since it was an area of great importance for RS development.

This paper is organized as what follows. In Section 2, the previous literature is reviewed. In the third section, we briefly define the terms used in this paper. In Section 4, the proposed model, which is a movie recommender system based on Firefly Algorithm (FA), K-Means, K-Nearest Neighbor (KNN) classification, and Bayesian classification, is presented. Section 5 is the experimental analysis using a MovieLens dataset, showing the results of the proposed approach. Finally, conclusion and the future works are presented.

## 2. Literature Review

The recommender systems help the users to find an appropriate option based on the user's personal taste and choice. They have been widely used to solve the information overload problem and thus have increased sales in e-commerce websites by understanding the customers' preferences and behavior [14]. Many RSs have been developed over the past decades, which usually use the data mining techniques in order to identify valid and useful patterns, and recommend the most appropriate suggestions to the users. To the best of the authors' knowledge, the previous studies used several methods to develop the recommender systems, which can be classified into the following categories: association rules, clustering, decision trees, K-Nearest Neighbors (KNN), link analysis, Artificial Neural Networks (ANN), Regression, and Heuristic methods, as discussed in Table 1. Also, the usage frequency of these methods is shown in Figure 1, which is drawn based on Table 1.

As it can be seen in Figure 1, the Heuristic method stands first in the body literature, which is due to the flexible nature of RSs. KNN is the second most significant method in the literature due to finding the most similar users to the underlying one. Clustering has the third most comprehensive share in the related literature.

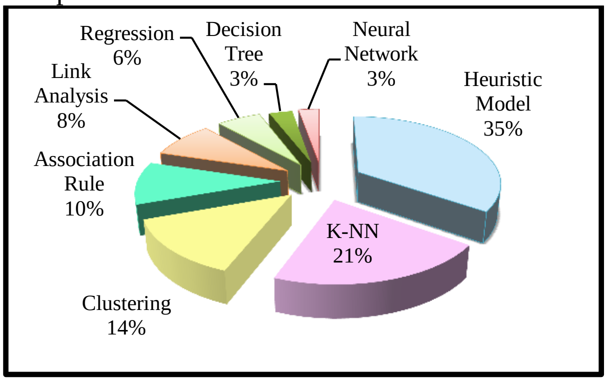


**Figure 1. Usage frequency of data mining techniques for recommendation.**

There has been a considerable interest in the hybrid approach in the recent years due to its effectiveness over the traditional approach [15, 16]. As the hybrid systems are a combination of multiple recommendations, alleviate the drawback of individual technique. Kumar *et al*. [17] have proposed a hybrid RS by combining content-based filtering and collaborative filtering. They used sentiment analysis in order to boost up the proposed RS. Pérez-Marcos *et al*. [16] have presented a hybrid system of video game recommendation through collaborative filtering and content-based filtering and the construction of relationship graphs to consider the hours of play. Harakawa *et al*. [18] have proposed a multi-modal Field-aware Factorization Machines (FFMs) algorithm to recommend the sentiment-aware personalized tweet. Since the sentiment factors strongly influence the users' interest in the tweet, this method models the users' interest by deriving multi-modal FFM that enables a collaborative use of the multiple elements in a tweet, and improves performance. Another study has developed a framework based on the user recommender interaction that takes input from the user, recommends N items to the user, and records the user choice until none of the recommended items favor [19]. The researchers have widely used the KNN approach in the recommender systems due to its efficiency, robustness, and interpretability [20, 21]. Pawar *et al*. [22] have designed a tour guide system based on three layers of architecture including the browser layer, the top layer, and the bottom layer. They used the KNN algorithm and collaborative filtering in order to calculate and recommend the tourism information to the users. Zhou and Yu [23] have developed a KNN classifier-based ensemble framework. Domeniconi and Yan [24] have studied the KNN ensemble approach and their relationship according to error correlation and accuracy. Argentini and Blanzieri [25] have suggested the neighborhood counting measure by considering the similarity measure of the KNN algorithm. Derrac *et al*. [26] have improved the performance of KNN by adopting the Cooperative co-evolution method.

Clustering is another approach that is occasionally used for recommendation. K-Means [27], K-Medoids [28], Self-Organizing Maps (SOM) [29], fuzzy C-Means [30], Expectation-Maximization (EM) [31] and the hierarchical techniques [32] have been applied to the RS problem. Cintia Ganesha Putri *et al*. [4] have developed the recommender system development using several algorithms to obtain groupings such as the K-Means algorithm, birch algorithm, mini-batch K-Means algorithm, mean-shift algorithm, affinity propagation algorithm, agglomerative clustering algorithm, and spectral clustering algorithm. In order to verify the recommender system's quality, they adopted the Mean Square Error (MSE) such as the Dunn Matrix and Cluster Validity Indices, and Social Network Analysis (SNA) such as Degree Centrality, Closeness Centrality, and Betweenness Centrality. Besides, two types of users' data are usually used for recommendation: demographic data and previous behavior patterns [33]. The present RSs are mostly or entirely on the users' previous behavior. In this research work, we also used the same data. Since applying the evolutionary algorithms and metaheuristics for clustering in RS's area has the potential to be effective, in this research work, we will propose a hybrid approach based on the firefly algorithm as a relatively new and robust evolutionary approach.

**Table 1. Classification of studies in RS's area based on the applied techniques in movie industry.**

| Author | KNN | Rule mining | Decision tree | clustering | Regression | Heuristics | Neural networks |
|---|---|---|---|---|---|---|---|
| Ramakrishnan ***et al***. [34] | ✓ | | | | | | |
| Herlocker and Konstan [35] | | ✓ | | | | | |
| Cheung ***et al***. [36] | ✓ | ✓ | ✓ | | | | |
| Roh ***et al***. [29] | ✓ | | | ✓ | | | |
| Cheung ***et al***. [37] | | | | ✓ | | | |
| Han ***et al***. [38] | ✓ | | | | | | |
| Weng and Liu [39] | ✓ | | | ✓ | | | |
| Zeng ***et al***. [40] | ✓ | | | | | | |
| Herlocker ***et al***. [41] | ✓ | | | | | | |
| Miller ***et al***. [42] | | ✓ | | | | | |
| Min and Han [43] | ✓ | | | ✓ | | | |
| Li ***et al***. [44] | ✓ | | | | | | |
| Kim and Yum [45] | | | | ✓ | | | |
| Lee ***et al***. [46] | | | | | ✓ | | |
| Salter and Antonopoulos [47] | | | | | | ✓ | |
| Du Boucher-Ryan and Bridge [48] | ✓ | ✓ | | | | | |
| Prangl ***et al***. [49] | | | | | | ✓ | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Hurley ***et al***. [50] | ✓ | | | | | |
| Im and Hars [51] | | | | | ✓ | |
| Symeonidis ***et al***. [52] | ✓ | | ✓ | | | |
| Chen ***et al***. [53] | ✓ | ✓ | | | | |
| Russell and Yoon [54] | | | | | ✓ | |
| Lee and Olafsson [55] | ✓ | | | | | |
| Jeong ***et al***. [56] | ✓ | | | | | |
| Jeong ***et al***. [57] | ✓ | | | | | |
| Acilar and Arslan [58] | ✓ | | ✓ | | | |
| Koren ***et al***. [59] | ✓ | | | | | |
| Chen ***et al***. [60] | ✓ | | | | | |
| Cho ***et al***. [61] | | | | | ✓ | |
| Bobadilla ***et al***. [62] | ✓ | | | | | |
| Julia ***et al***. [63] | | | | | ✓ | |
| Winoto and Tang [64] | | | | | ✓ | |
| Ahn ***et al***. [65] | | | | | ✓ | |
| Bobadilla ***et al***. [66] | ✓ | | | | | |
| Ozok ***et al***. [67] | | | | ✓ | | |
| Huang [68] | | ✓ | | ✓ | ✓ | |
| Ghazanfar and Prügel-Bannett [69] | | | ✓ | ✓ | ✓ | |
| Li ***et al***. [70] | | | ✓ | | | |
| Liang ***et al***. [71] | | | | | ✓ | |
| Liu ***et al***. [72] | | | | | ✓ | |
| Sun ***et al***. [73] | | | ✓ | | | |
| Zahra ***et al***. [74] | | | | | ✓ | |
| Koohi and Kiani [30] | | | ✓ | | | |
| Hernando ***et al***. [75] | | | | | ✓ | |
| Da Silva ***et al***. [76] | | | | | ✓ | |
| Ar and Bostanci [77] | | | | | ✓ | |
| Nilashi ***et al***. [78] | | | | | ✓ | |
| Liao and Lee [79] | | | ✓ | | | |
| Kermany and Alizadeh [80] | | | | | ✓ | |
| Ebesu and Fang [81] | | | | | ✓ | |
| Koohi and Kiani [82] | | | | | ✓ | |
| Park and Kim [83] | | | ✓ | | | |
| Najafabadi ***et al***. [84] | | ✓ | ✓ | | | |
| Jomsri [85] | | ✓ | | | | |
| Li ***et al***. [86] | | | | ✓ | | |
| Chen ***et al***. [87] | | ✓ | ✓ | | | |
| Tewari and Barman [88] | | | | | ✓ | |
| Viktoratos ***et al***. [89] | | | | | ✓ | |
| Lin and Chi [12] | | | | | | ✓ |
| Rajarajeswari ***et al***. [1] | | | | | ✓ | |
| Ahuja ***et al***. [90] | ✓ | | ✓ | | | |
| Cintia Ganesha Putri ***et al***. [4] | | | ✓ | | | |
| Kumar ***et a***l. [17] | | | | | ✓ | |
| Pérez-Marcos ***et al***. [16] | | | | | ✓ | |
| Hassanpour ***et al***. [91] | | | | | ✓ | |

## 3. Preliminaries

In this research work, the proposed RS is based on the clustering techniques (K-Means), evolutionary algorithms (Firefly algorithm), and classification techniques (KNN and Naïve Bayesian classification). Here, we will have a quick review of these techniques.

### 3.1. K-Means Clustering

Clustering is an analytical method used as early as 1939 by Tryon, R. C, defined as the grouping of objects into relatively homogeneous sub-groups or clusters [92]. *K*-Means clustering is a method that automatically divides the datasets into *k* groups. *K*-Means is used to group the approaches given its simplicity, efficiency, and flexibility in calculations, especially considering large data. The procedure of the K-means algorithm is as follows [93]:

1- Specify the number of clusters, and randomly select *k* objects as the initial center of these clusters.

2- For each remaining object, assign it to the cluster whose center is the nearest.

3- Find the cluster centroid for each cluster, and assign the centroid as the cluster's new center.

4- Repeat steps 2 and 3 until the center of clusters is fixed or other predefined termination conditions are satisfied.

### 3.2. Firefly Algorithm (FA)

FA is a population-based optimization algorithm, and mimics a firefly's attraction to flashing light. This algorithm has been proposed by Yang [94], which is classified as swarm intelligent, metaheuristic, and nature-inspired. This algorithm is naturally a multi-modal algorithm. Therefore, it can be suitable for the structural engineering problems, especially when we need to prepare the engineering alternatives in multi-modal issues. Furthermore, the method's effectiveness was validated compared with the artificial neural networks, genetic programming models, and Particle Swarm Optimization (PSO) [94, 95].

The flashing light of fireflies is a way to attract mating partners as well as attracting potential prey. This behavior would be mimicked to solve the optimization problems. The basic FA algorithm can be described as follows. The light intensity $I$ decrease in terms of $I \propto 1/r^2$, in which $r$ is the light source's distance. Moreover, the light becomes weaker as the distance increases due to air absorption [94].

Three primary rules are established for this algorithm [94]:

- All fireflies are unisex. Therefore, sex is not a matter of importance in attracting other fireflies.
- The attractiveness of a firefly is proportional to its brightness intensity.
- The brightness of a firefly is affected or determined by the objective function.

Here, the light intensity $I$ is proportional to the objective function of the optimization model; in other words, $I(s) \propto F(s)$, in which $s$ is a solution or a location. The brightness or light intensity $I$ or attractiveness are the same in FA, which is relative and varies with the distance between the source and destination [94]. Therefore, the light intensity $I(r)$ can be measured as (1):

$$I(r) = I_0 e^{-\gamma r^2} \quad (1)$$

where $I_0$ is the initial light intensity at the source, and $\gamma$ represents a fixed light absorption coefficient. Similarly, the attractiveness $\beta$ is defined by (2):

$$\beta(r) = \beta_0 e^{-\gamma r^2} \quad (2)$$

where $\beta_0$ is attractiveness at the source. In order to obtain $r$ in (1) and (2), we know that the distance between the two fireflies $i$ and $j$ is calculated as (3):

$$r_{ij} = \| S_i - S_j \| = \sqrt{\sum_{K=1}^{D} (S_{ik} - S_{jk})^2} \quad (3)$$

in which $S_{ik}$ and $S_{jk}$ are the location of the $i$th and $j$th fireflies, respectively. $D$ refers to the Cartesian space dimension, and $k$ is a numerator between 1 and $D$.

Each firefly $i$ is attracted to a firefly $j$ by (4) as follows:

$$S_i = S_i + \beta_0 e^{-\gamma r_{ij}^2} (S_j - S_i) + \alpha N_i(0,1) \quad (4)$$

The first term is the firefly's current location, and the second term refers to the attractiveness. The third term with $\alpha$ is the randomization, and $N_i(0,1)$ is a normal random variable.

Consider a population of fireflies $P^{(t)}$ with the vector of locations: $S_i^{(t)} = S_{i0}^{(t)}, \ldots, S_{in}^{(t)}$, which $I=1, \ldots, NP$ and $NP$ refers to the number of fireflies in the population $P^{(t)}$ in the $t$th generation. Equation (5) determines the location of the initial population:

$$S_{ij}^{(0)} = (ub_i - lb_i).rand(0,1) + lb_i \quad (5)$$

where $ub_i$ and $lb_i$ are the upper and lower bounds, respectively.

The firefly search process comprises the following steps:

1. At first, the initial value of $\alpha$ (randomization variable) is updated based on (6) and (7):

$$\Delta = 1 - 10^{-4} / 0.9^{1/MAX\_GEN} \quad (6)$$

$$\alpha^{(t+1)} = 1 - \Delta.\alpha^{(t)} \quad (7)$$

Where Δ obtains the step size, which decreases with increases in *t* (number of generations).

2. The new $S_i^{(t)}$ is evaluated based on $F(S^t)$, in which $S(x_i^{(t)}) = S_i^{(t)}$.
3. In this step, the algorithm sorts the population according to their $F(S^t)$ in increasing order.
4. The algorithm finds the best individual in $P^{(t)}$, which is represented by $S^*$.
5. Finally, according to their attractiveness, the fireflies move to their neighbor fireflies in the search space.

A pre-defined maximum number of evaluations usually controls the firefly search process.

### 3.3. K-nearest Neighbor Classification

The *k*-nearest-neighbor approach to classification is relatively simple, completely nonparametric. A user must make only two choices: (1) the number of neighbors, *k*, and (2) the distance metric to be used. The common options of distance metrics include the Euclidean distance, Mahalanobis distance, and city-block distance. The number of neighbors is usually selected by either cross-validation or testing the classifier's quality on a second test data set [96]. KNN is an instance-based supervised learning method popular for the recommendation. The following steps are usually necessary for a recommender system [97]:

1. Make a profile for each user based on their preferences.
2. Find the *k* most similar or nearest neighbors to the target user.
3. Identify the most *n* special items that the target user may purchase based on the neighbor's preferences.

### 3.4. Bayesian Classification

Bayesian networks are one of the most promising approaches in data mining, which are a sub-class from the probabilistic graphical models [98]. Bayesian networks are used in the recommendation process as a standalone model or used parallel with the other techniques [99]. It aims to apply both the graph theory and the probability theory in order to solve complex problems by decomposing them into small elements. A graph-based probabilistic model has been defined where the nodes represent the random variables, and the edges of the graph show the dependencies among the variables. The conditional probabilities giving the distribution over the variables are shown in the Conditional Probability Table (CPTs). A Naive Bayes classifier is a simple probabilistic classifier based on applying the Bayes theorem (from Bayesian statistics) with (naive) independence assumptions. An advantage of the naive Bayes classifier is that it only requires a small amount of training data in order to estimate the classification parameters [98].

## 4. Proposed Method

In this sub-section, we introduce our FA K-Means algorithm based on the Firefly Algorithm and K-Means combination.

### 4.1. Proposed FA K-Means Clustering Algorithm

Although FA is useful in searching for the feasible region, it is not as good as K-Means in fine-tuning. However, K-Means lacks the ability to a global search [93]. In the local search methods, i.e. K-Means, the final solution greatly depends on the initial solution so our proposed search method escapes from local solutions considering FA. It seems that combining these two techniques would be useful in our clustering application. Therefore, in this research work, we use an FA K-Means algorithm. This algorithm first provides the solutions based on FA, and then K-Means is implemented to find the centroids. The proposed algorithm includes the following steps:

Step 1. Determine the values for the following parameters: $\alpha, \beta, k, \gamma$.

Step 2. Set counter *maxit = 0*.

Step 3. Determine the values of the initial population. The FA algorithm considers the lower and upper bounds for the values, and then a random initial population is produced.

Step 4. Apply the FA algorithm based on the following sub-steps:

*4.1.* Apply the following tasks in order to update the population: Choose an appropriate fitness function *F(s)* to evaluate the solutions' quality. Select the fitness function as ($X+a$), in which $X$ is the location of each firefly (the initial location is randomly determined), and $a$ is a constant obtained by trial-and-error.

*4.2.* Update the number of iterations. If *maxit < 40*, go to *4-1*; else go to the next step.

Step 5. Use the output of FA to obtain the centroid of the clusters to K-Means. The following sub-steps will be done:

*5.1.* Run K-Means for updating the clusters and finding their centroids.

*5.2.* Update the number of iterations.

Continue the process until the termination conditions are satisfied.

### 4.2. Proposed Recommender System

We implement our proposed recommender system in two stages (Figure 2). In the first stage, the objects will be mapped into the specified *k* clusters using the FA K-Means algorithm. The number of clusters is obtained by tuning for different values of k using the Davies-Bouldin index. In the next stage, the new user (target customer) either enters his/her demographic data such as age, gender, job or the data would be extracted from a profile. The system then utilizes the past demographic characteristics to assign the active user to the nearest cluster, which embodies similar information. Finally, the Naïve Bayesian and KNN will be used for recommendation.

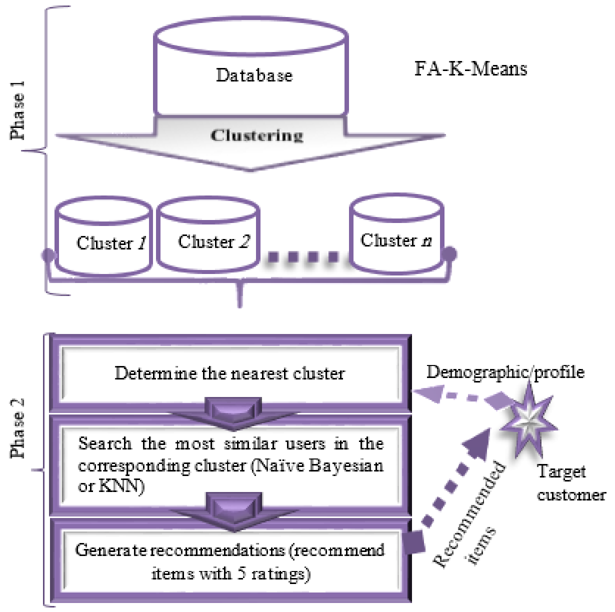


**Figure 2. Proposed framework for recommendation.**

### 4.3. Model Evaluation

In order to ensure that the proposed model gives the best accuracy in the recommendation system, we applied different evaluation measures. We evaluated the proposed FA K-Means clustering algorithm using Davies-Bouldin (DB), an appropriate index to measure clustering validity. Moreover, accuracy, precision, and recall were used to evaluate the proposed RS.

#### 4.3.1. Davies-Bouldin (DB) Index

The DB index is an internal scheme that evaluates the effect of clustering based on the dataset's scale and characteristics. The basic idea is to measure the clustering's impact by calculating the within-cluster similarity ratio and among-cluster similarity. Since the clustering method's vision is to make the within-cluster similarity as large as possible and otherwise the among-cluster resemblance, it is evident that the smaller the DB index, the better the clustering result [100]. The scatter within *the* *i*th cluster, and the distance between the *i*th and *j*th clusters are calculated as (8) and (9):

$$S_i = \left[\frac{1}{N_i}\sum_{x\in C_i}\|X - m_i\|_2^q\right]^{1/q} \tag{8}$$

$$d_{i,j} = \|\vec{m}_i - \vec{m}_j\|_t = \left\{\sum_{p=1}^{d}|m_{i,p} - m_{j,p}|^t\right\}^{\frac{1}{t}} \tag{9}$$

where $m_i$ is the centroid of the *ith* cluster, and *q* and *t* are integers greater or equal to 1. They can be selected independently. $N_i$ represents the number of objects in the *i*th cluster [100]. Then $R_{i,q}$ is defined as (10):

$$R_{i,q} = \max_{j\in K, j\neq i}\left\{\frac{S_{i,q}+S_{j,q}}{d_{ij}}\right\} \tag{10}$$

Finally, based on the above definitions, DB is defined as (11):

$$DB(K) = \frac{1}{k}\sum_{i=1}^{K} R_{i,q} \tag{11}$$

#### 4.3.2 Measures of RS Performance

In order to evaluate the generated RS's performance, we used four criteria evaluation that were accuracy, precision, recall, and F-measure. Defining the criteria mentioned above, we should first introduce a confusion matrix (Table 2). This matrix provides information about the actual and predicted classifications of a classification algorithm [101].

**Table 2. Confusion matrix [101].**

| | | Predicted result | |
|---|---|---|---|
| **Actual result** | | Recommend | Not recommend |
| | Recommend | True positives (TP) | False positives (FP) |
| | Not recommend | False negatives (FN) | True negatives (TN) |

According to the confusion matrix, the measures are defined as (12), (13), (14), and (15) [101]:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{12}$$

$$Precision = \frac{TP}{TP+FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$F - measure = \frac{2 \times recall \times precision}{recall + precision} \quad (15)$$

## 5. Experiment and Analysis

In this section, the proposed recommendation algorithms were experimentally evaluated for the recommendations' efficiency, performance, and effectiveness with the existing approaches. These experiments were performed in MATLAB on Intel(R) Core(TM) i5-2400 CPU @ 3.10 GHz, 8.0 GB RAM computer.

### 5.1. Dataset and Data Pre-processing

The MovieLens datasets are considered as the standard datasets in evaluating the recommendation approaches; the dataset was taken from the Kaggle website. This dataset comprises the users' ratings and appropriate demographic information. The MovieLens ratings are collected on the 5-point rating scale; the 5-star rating is highly liked, and the 1-star rating is most hated or disliked. The dataset consists of data from 943 users about 1682 movies with 100,000 ratings [102].

We used this dataset in order to evaluate our proposed recommendation approach. The dataset consists of the users' demographic information such as age, gender, zip code, and occupation; in this work, we used age, gender, and occupation for further processing. Besides, since the similarity criterion used in this work was the Euclidean distance, it was necessary to normalize the data used (age, gender, and occupation). For this purpose, the definition of binary variables was used for each character so that it was one bit for gender, four bits for the job, and six bits for age (age between 15 to 79 years). Having done the data pre-processing operation, the researchers can use this dataset readily. This pre-processing was done by deleting all the users who did not enter their demographic information or rated less than twenty videos. As a result, there is no need to trim and pre-process the data to use the dataset mentioned above. Table 3 shows some of the characteristics of the dataset.

**Table 3. Main characteristics of selected dataset [102].**

| Name | Number of ratings (1-5) | Number of users | Number of movies | Demographic attributes |
|---|---|---|---|---|
| MovieLens | 100,000 | 943 | 1682 | Age, gender, occupation |

### 5.2. Algorithm Clustering Result

We tried to establish a novel clustering algorithm using FA and K-Means in order to discover the similarities within the groups of people to develop a movie recommendation system for the users. According to the literature, we used the Euclidean distance to measure the similarity between the objects to form the clusters. Hence, we normalized the values for the mentioned attributes. We defined a binary variable to each one of the following attributes: one-bit variable for gender, four-bit variable for occupation, and six-bit variable for age (varied between 15 and 79). Table 4 shows the optimal values for the parameters that are gained after tuning.

**Table 4. Optimal parameters for FA K-Means approach.**

| Initial population | α | β | γ | k |
|---|---|---|---|---|
| 60 | 0.1 | 0.5 | 1 | 10 |

Comparing the results obtained from the FA K-Means algorithm shows a significant improvement than that of K-Means. Table 5 shows this remarkable improvement based on the previously introduced DB index. The costs are obtained after running 50 iterations of the proposed algorithm.

**Table 5. Results of 50 iterations of the proposed algorithm.**

| Algorithm | Cost function | | |
|---|---|---|---|
| | Min cost | Average cost | Max cost |
| FA K-Means | 46 | 103 | 117 |

In Table 6, the results of the DB index measurement are presented. The results obtained show an impressive improvement by combining FA with K-Means.

**Table 6. Comparison of the proposed model with K-Means.**

| Algorithm | DB value |
|---|---|
| K-Means | 0.2515 |
| FA-K-Means | 0.1352 |

Figures (3) to (5) show the results of applying K-Means, FA, and FA K-Means. These figures' vertical axis offers the best cost or total distances, obtained through (4) and (16). Horizontal axes show the iterations. The figures illustrate the proposed algorithm's improvement inseparability. Figure (5) shows that the similarity index among intra-cluster objects is improved, and the dissimilarity index among inter-cluster objects is compared with the results of the previous algorithm (See Figure 3). These improvements have appeared in the DB index.

$$Best\cos t = \min \sum_{i=1}^{npop} S_i \quad (16)$$

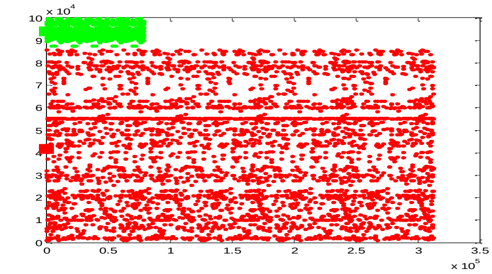

**Figure 3. Output of K-Means.**

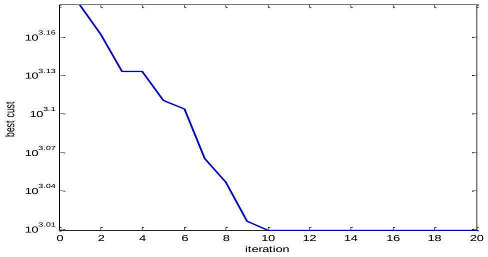


**Figure 4. Output of FA.**

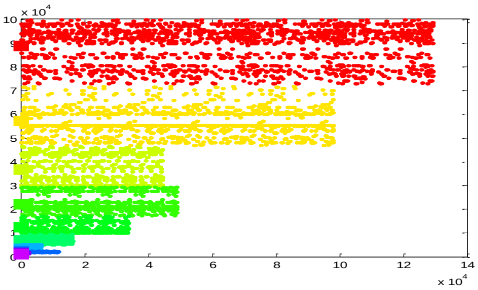

**Figure 5. Output of FA K-Means.**

### 5.3. Analysis of Results

The proposed recommender system results based on accuracy, precision, and recall rates are shown in Table 7. The performance evaluation showed that KNN with k = 5 (5NN) and a recommendation with rank 5 provides a good performance.

The achieved accuracy, precision, recall, and F-measure are considered significant compared to the previous literature findings.

Naïve Bayesian classification, the other applied technique for the recommendation, also achieved better results than the previous findings. Comparing KNN and Naïve Bayesian, KNN outperforms Naïve Bayesian in accuracy and recall; however, it is not practical in terms of precision. Due to the nature of the problem and data, the accuracy and recall rates seem more significant. Thus, KNN is more effective in this case. For more clarification, the confusion matrix is presented in Table 8.

**Table 7. Comparison of techniques used in this research work and previous studies.**

| Research | Used technique | Accuracy | Recall | Precision | F-measure |
|---|---|---|---|---|---|
| Proposed method | K-NN | 78.31% | 54.57% | 93.52% | 68.92% |
| | Naïve Bayesian | 58.69% | 24.1% | 98.48% | 38.72% |
| Tsai and Hung [103] | SOM Ensembles | 77.09% | 20.68% | 76.24% | 32.53% |
| | K-Means Ensembles | 77.86% | 22.91% | 75.64% | 35.17% |
| Renaud-Deputter *et al*. [104] | Global Top-N | - | - | 66.82% | - |
| | Weighted user-based K-NN | - | - | 74.28% | - |
| | Weighted item-based | - | - | 72.75% | - |
| | WR-MF | - | - | 73.57% | - |
| | BPR-MF | - | - | 72.64% | - |
| | Their system | - | - | 74.75% | - |

**Table 8. Confusion matrix for the current problem.**

| | | Predicted values (recommended) | |
|---|---|---|---|
| **Actual results (user's actual interests)** | | OK (cluster users have given a high rating to the movie so that the movie will be recommended) | Not OK (cluster users have given a low rating to the movie so that the movie will not be recommended) |
| | Like (user is, in fact, part of the cluster where the users like the movie) | LO.1 | N.2 |
| | Dislike (user is, in fact, part of the cluster where the users do not like the movie) | DO.3 | NO.4 |

## 6. Conclusions

The recommender systems have attracted the attention of many researchers in the recent years. Still, several limitations and shortcomings such as a gap between the models' expectations and performance are required to be addressed. In this work, we developed a recommendation system in order to address these problems. We proposed a new movie recommender in which the k-Means and firefly algorithms were applied to the cluster objects before using KNN and Naïve Bayesian classification in order to find similar objects for the recommendation. As the MovieLens datasets are considered the standard datasets in evaluating the recommendation approaches, in this research work, a dataset from the MovieLens was selected and used for model training and test. Both the demographic characteristics of the users and their rankings are available in the dataset.

In order to find the optimal clustering based on FA and k-Means, DBI was used, which proved a significant improvement compared with K-Means. Then KNN and Naïve Bayesian classification were applied to provide recommendations for the new users. Accuracy, precision, recall, and F-measure were calculated for each technique. The results obtained show the superiority of KNN over Naïve Bayesian in terms of accuracy and recall. On the other hand, Naïve Bayesian offers better results compared with KNN in terms of precision. Since, in this research work, accuracy and recall are a matter of concern, KNN seems useful for movie recommendations. The parameter tuning shows that K = 5 provides better results, and therefore, to recommend a movie, the five most similar users should be identified and considered. Not only the proposed approach shows its applicability and effectiveness for movie recommendations but also the other products and services can benefit from this approach.

For future works, the proposed algorithm will be tested in a different kind of dataset. Moreover, the proposed algorithms will be implemented by considering a dataset containing other factors such as social or trust relationships using Epinion dataset, for example. The Epinion dataset is a product review website that started in 1999 (www.epinion.com). On this website, the users can rate items from 1 to 5 and submit their personal reviews. The users can also express their web of trust. Then we can use the MovieLense dataset in order to evaluate the recommendation algorithms.

## References

[1] S. Rajarajeswari, S. Naik, S. Srikant, M. S. Prakash, and P. Uday, "Movie Recommendation System," In *Emerging Research in Computing, Information, Communication and Applications.* Springer, Singapore, pp. 329-340, 2019.

[2] X. Yang, Y. Guo, Y. Liu, and H. Steck, "A survey of collaborative filtering based social recommender systems," Computer communications, Vol. 41, pp. 1-10, 2014.

[3] H. Tahmasebi, R. Ravanmehr, and R. Mohamadrezaei, "Social movie recommender system based on deep autoencoder network using Twitter data," *Neural Computing and Applications*, pp. 1-17, 2020.

[4] D. Cintia Ganesha Putri, J. S. Leu, and P. Seda, "Design of an Unsupervised Machine Learning-Based Movie Recommender System," *Symmetry*, Vol. 12, No. 2, pp. 185, 2020.

[5] L. Esmaeili, S. Mardani, S. A. H. Golpayegani, and Z. Z. Madar, "A novel tourism recommender system in the context of social commerce," *Expert Systems with Applications*, Vol. 149, pp. 113301, 2020.

[6] T. N. T. Tran, M. Atas, A. Felfernig, and M. Stettinger, "An overview of recommender systems in the healthy food domain," *Journal of Intelligent Information Systems*, Vol. 50, No. 3, pp. 501-526, 2018.

[7] S. R. S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, and B. Venkatesh, "Content-based movie recommendation system using genre correlation," *In Smart Intelligent Computing and Applications* Springer, Singapore, pp. 391-397, 2019.

[8] Q. Li, I. Choi, and J. Kim, Evaluation of Recommendation System for Sustainable E-Commerce: Accuracy, Diversity and Customer Satisfaction, 2020.

[9] G. Desirena, A. Diaz, J. Desirena, I. Moreno, and D. Garcia, "Maximizing Customer Lifetime Value using Stacked Neural Networks: An Insurance Industry Application," *in 18th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 2019, pp. 541-544.

[10] Y. Bai, S. Jia, S. Wang, and B. Tan, "Customer Loyalty Improves the Effectiveness of Recommender Systems Based on Complex Network," *Information*, Vol. 11, No. 3, pp. 171, 2020.

[11] P. Vilakone, K. Xinchang, and D. S. Park, "Movie recommendation system based on users' personal information and movies rated using the method of k-clique and normalized discounted cumulative gain," *Journal of Information Processing Systems*, Vol. 16, No. 2, pp. 494-507, 2020.

[12] C. H. Lin, and H. Chi, "A novel movie recommendation system based on collaborative filtering and neural networks," In *International*

*Conference on Advanced Information Networking and Applications*. Springer, Cham, 2019, pp. 895-903.

[13] M. Y. Hsieh, W. K. Chou, and K. C. Li, "Building a mobile movie recommendation service by user rating and APP usage with linked data on Hadoop," *Multimedia Tools and Applications*, Vol. 76, No. 3, pp. 3383-3401, 2017.

[14] M. Ashraf, S. Ouf, and Y. Helmy, "A Proposed Paradigm for Enhancing Customer Retention using Web Usage Mining," *International Journal of Computer Applications*, Vol. 975, pp. 8887, 2020.

[15] M. Riyahi, and M. K. Sohrabi, "Providing effective recommendations in discussion groups using a new hybrid recommender system based on implicit ratings and semantic similarity," *Electronic Commerce Research and Applications*, Vol. 40, pp. 100938, 2020.

[16] J. Pérez-Marcos, L. Martín-Gómez, D. M. Jiménez-Bravo, V. F. López, and M. N. Moreno-García, "Hybrid system for video game recommendation based on implicit ratings and social networks," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-11, 2020.

[17] S. Kumar, K. De, and P. P. Roy, "Movie recommendation system using sentiment analysis from microblogging data," *IEEE Transactions on Computational Social Systems*, 2020.

[18] R. Harakawa, D. Takehara, T. Ogawa, and M. Haseyama, "Sentiment-aware personalized tweet recommendation through multimodal FFM," *Multimedia Tools and Applications*, Vol. 77, No. 14, pp. 18741-18759, 2018.

[19] H. R. Zhang, F. Min, X. He, and Y. Y. Xu, "A hybrid recommender system based on user-recommender interaction," *Mathematical Problems in Engineering*, 2015.

[20] M. Ludewig, I. Kamehkhosh, N. Landia, and D. Jannach, "Effective nearest-neighbor music recommendations," In *Proceedings of the ACM Recommender Systems Challenge*, pp. 1-6, 2018.

[21] M. Ludewig, and D. Jannach, "Evaluation of session-based recommendation algorithms," *User Modeling and User-Adapted Interaction*, Vol. 28, No. 4-5, pp. 331-390, 2018.

[22] S. S. Pawar, A. S. Kadan, P. R. Chavhan, P. R. Ranjane, and A. S. Lohar, "Android Based Tourist Guide System," *International Journal of Engineering Technology, Management and Applied Sciences (IJETMAS)*, Vol. 4, No. 2, pp. 42-46, 2016.

[23] Z. H. Zhou, and Y. Yu, "Ensembling local learners Through Multimodal perturbation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 35, No. 4, pp. 725-735, 2005.

[24] C. Domeniconi, and B. Yan, "Nearest neighbor ensemble," In *Proceedings of the 17th International Conference on Pattern Recognition*, IEEE, Vol. 1, 2004, pp. 228-231.

[25] A. Argentini, and E. Blanzieri, "About neighborhood counting measure metric and minimum risk metric," *IEEE transactions on pattern analysis and machine intelligence*, Vol. 32, No. 4, pp. 763-765, 2009.

[26] J. Derrac, I. Triguero, S. García, and F. Herrera, "Integrating instance selection, instance weighting, and feature weighting for nearest neighbor classifiers by coevolutionary algorithms," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 42, No. 5, pp. 1383-1397, 2012.

[27] S. Rajabi, A. Harounabadi, and V. Aghazarian, "A recommender system for the web: using user profiles and machine learning methods," *International Journal of Computer Applications*, Vol. 96, No. 11, 2014.

[28] G. Guo, J. Zhang, and N. Yorke-Smith, "Trustsvd: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings," In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29, No. 1, 2015.

[29] T. H. Roh, K. J. Oh, and I. Han, "The collaborative filtering recommendation based on SOM cluster-indexing CBR," *Expert systems with applications*, Vol. 25, No. 3, pp. 413-423, 2003.

[30] H. Koohi, and K. Kiani, "User-based Collaborative Filtering using fuzzy C-means," *Measurement*, Vol. 91, pp. 134-139, 2016.

[31] A. R. Anaya, M. Luque, and T. García-Saiz, "Recommender system in collaborative learning environment using an influence diagram," *Expert Systems with Applications*, Vol. 40, No. 18, pp. 7193-7202, 2013.

[32] Y. Wang, W. Dai, and Y. Yuan, "Website browsing aid: A navigation graph-based recommendation system," *Decision support systems*, Vol. 45, No. 3, pp. 387-400, 2008.

[33] J. B. Schafer, J. Konstan, and J. Riedl, "Recommender systems in e-commerce," In *Proceedings of the 1st ACM conference on Electronic commerce*, 1999, pp. 158-166.

[34] N. Ramakrishnan, B. J. Keller, B. J. Mirza, A. Y. Grama, and G. Karypis, "When being weak is brave: Privacy in recommender systems," *arXiv preprint cs/0105028*, 2001.

[35] J. L. Herlocker, and J. A. Konstan, "Content-independent task-focused recommendation," *IEEE Internet Computing*, Vol. 5, No. 6, pp. 40-47, 2001.

[36] K. W. Cheung, J. T. Kwok, M. H. Law, and K. C. Tsui, "Mining customer product ratings for personalized marketing," *Decision Support Systems*, Vol. 35, No. 2, pp. 231-243, 2003.

[37] K. W. Cheung, K. C. Tsui, and J. Liu, "Extended latent class models for collaborative recommendation," *IEEE Transactions on Systems, Man, and Cybernetics-*

*Part A: Systems and Humans*, Vol. 34, No. 1, pp. 143-148, 2004.

[38] P. Han, B. Xie, F. Yang, and R. Shen, “A scalable P2P recommender system based on distributed collaborative filtering,” *Expert systems with applications*, Vol. 27, No. 2, pp. 203-210, 2004.

[39] S. S. Weng, and M. J. Liu, “Feature-based recommendations for one-to-one marketing,” *Expert Systems with Applications*, Vol. 26, No. 4, pp. 493-508, 2004.

[40] C. Zeng, C. X. Xing, L. Z. Zhou, and X. H. Zheng, “Similarity measure and instance selection for collaborative filtering,” *International Journal of Electronic Commerce*, Vol. 8, No. 4, pp. 115-129, 2004.

[41] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, “Evaluating collaborative filtering recommender systems,” *ACM Transactions on Information Systems (TOIS)*, Vol. 22, No. 1, pp. 5-53, 2004.

[42] B. N. Miller, J. A. Konstan, and J. Riedl, “PocketLens: Toward a personal recommender system,” *ACM Transactions on Information Systems (TOIS)*, Vol. 22, No. 3, pp. 437-476, 2004.

[43] S. H. Min, and I. Han, “Detection of the customer time-variant pattern for improving recommender systems,” *Expert Systems with Applications*, Vol. 28, No. 2, pp. 189-199, 2005.

[44] Y. Li, L. Lu, and L. Xuefeng, “A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce,” *Expert systems with applications*, Vol. 28, No. 1, pp. 67-77, 2005.

[45] D. Kim, and B. J. Yum, “Collaborative filtering based on iterative principal component analysis,” *Expert Systems with Applications*, Vol. 28, No. 4, pp. 823-830, 2005.

[46] J. S. Lee, C. H. Jun, J. Lee, and S. Kim, “Classification-based collaborative filtering using market basket data,” *Expert systems with applications*, Vol. 29, No. 3, pp. 700-704, 2005.

[47] J. Salter, and N. Antonopoulos, “CinemaScreen recommender agent: combining collaborative and content-based filtering,” *IEEE Intelligent Systems*, Vol. 21, No. 1, pp. 35-41, 2006.

[48] P. du Boucher-Ryan, and D. Bridge, “Collaborative recommending using formal concept analysis,” In *International Conference on Innovative Techniques and Applications of Artificial Intelligence.* Springer, London, 2005, pp. 205-218.

[49] M. Prangl, T. Szkaliczki, and H. Hellwagner, “A framework for utility-based multimedia adaptation,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 6, pp. 719-728, 2007.

[50] N. J. Hurley, M. P. O'Mahony, and G. C. Silvestre, “Attacking recommender systems: A cost-benefit analysis,” *IEEE Intelligent Systems*, Vol. 22, No. 3, pp. 64-68, 2007.

[51] I. Im, and A. Hars, “Does a one-size recommendation system fit all? effectiveness of collaborative filtering-based recommendation systems across different domains and search modes,” *ACM Transactions on Information Systems (TOIS)*, Vol. 26, No. 1, 2007.

[52] P. Symeonidis, A. Nanopoulos, A. N. Papadopoulos, and Y. Manolopoulos, “Collaborative recommender systems: Combining effectiveness and efficiency,” *Expert Systems with Applications*, Vol. 34, No. 4, pp. 2995-3013, 2008.

[53] Y. L. Chen, L. C. Cheng, and C. N. Chuang, “A group recommendation system with consideration of interactions among group members,” *Expert systems with applications*, Vol. 34, No. 3, pp. 2082-2090, 2008.

[54] S. Russell, and V. Yoon, “Applications of wavelet data reduction in a recommender system,” *Expert Systems with Applications*, Vol. 34, No. 4, pp. 2316-2325, 2008.

[55] J. S. Lee, and S. Olafsson, “Two-way cooperative prediction for collaborative filtering recommendations,” *Expert Systems with Applications*, Vol. 36, No. 3, pp. 5353-5361, 2009.

[56] B. Jeong, J. Lee, and H. Cho, “User credit-based collaborative filtering,” *Expert Systems with Applications*, Vol. 36, No. 3, pp. 7309-7312, 2009.

[57] B. Jeong, J. Lee, and H. Cho, “An iterative semi-explicit rating method for building collaborative recommender systems,” *Expert Systems with Applications*, Vol. 36, No. 3, pp. 6181-6186, 2009.

[58] A. M. Acilar, and A. Arslan, “A collaborative filtering method based on artificial immune network,” *Expert Systems with Applications*, Vol. 36, No. 4, pp. 8324-8332, 2009.

[59] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, Vol. 42, No. 8, pp. 30-37, 2009.

[60] G. Chen, F. Wang, and C. Zhang, “Collaborative filtering using orthogonal nonnegative matrix tri-factorization,” *Information Processing & Management*, Vol. 45, No. 3, pp. 368-379, 2009.

[61] J. Cho, K. Kwon, and Y. Park, “Q-rater: A collaborative reputation system based on source credibility theory,” *Expert Systems with Applications*, Vol. 36, No. 2, pp. 3751-3760, 2009.

[62] J. E. S. U. S. Bobadilla, F. Serradilla, and A. Hernando, “Collaborative filtering adapted to recommender systems of e-learning,” *Knowledge-Based Systems*, Vol. 22, No. 4, pp. 261-265, 2009.

[63] C. Julià, A. D. Sappa, F. Lumbreras, J. Serrat, and A. López, “Predicting missing ratings in recommender

systems: Adapted factorization approach," *International Journal of Electronic Commerce*, Vol. 14, No. 2, pp. 89-108, 2009.

[64] P. Winoto, and T. Y. Tang, "The role of user mood in movie recommendations," *Expert Systems with Applications*, Vol. 37, No. 8, pp. 6086-6092, 2010.

[65] H. J. Ahn, H. Kang, and J. Lee, "Selecting a small number of products for effective user profiling in collaborative filtering," *Expert Systems with Applications*, Vol. 37, No. 4, pp. 3055-3062, 2010.

[66] J. Bobadilla, F. Serradilla, and J. Bernal, "A new collaborative filtering metric that improves the behavior of recommender systems," *Knowledge-Based Systems*, Vol. 23, No. 6, pp. 520-528, 2010.

[67] A. A. Ozok, Q. Fan, and A. F. Norcio, "Design guidelines for effective recommender system interfaces based on a usability criteria conceptual model: results from a college student population," *Behavior and Information Technology*, Vol. 29, No. 1, pp. 57-83, 2010.

[68] S. L. Huang, "Designing utility-based recommender systems for e-commerce: Evaluation of preference-elicitation methods," *Electronic Commerce Research and Applications*, Vol. 10, No. 4, pp. 398-407, 2011.

[69] M. A. Ghazanfar, and A. Prügel-Bennett, "Leveraging clustering approaches to solve the gray-sheep user's problem in recommender systems," *Expert Systems with Applications*, Vol. 41, No. 7, pp. 3261-3275, 2014.

[70] X. Li, M. Wang, and T. P. Liang, "A multi-theoretical kernel-based approach to social network-based recommendation," *Decision Support Systems*, Vol. 65, pp. 95-104, 2014.

[71] W. Liang, G. Lu, X. Ji, J. Li, and D. Yuan, "Difference factor'KNN collaborative filtering recommendation algorithm," In *International Conference on Advanced Data Mining and Applications.* Springer, Cham, 2014, pp. 175-184.

[72] W. Liu, C. Wu, B. Feng, and J. Liu, "Conditional preference in recommender systems," *Expert Systems with Applications*, Vol. 42, No. 2, 774-788, 2015.

[73] Z. Sun, L. Han, W. Huang, X. Wang, X. Zeng, M. Wang, and H. Yan, "Recommender systems based on social networks," *Journal of Systems and Software*, Vol. 99, pp. 109-119, 2015.

[74] S. Zahra, M. A. Ghazanfar, A. Khalid, M. A. Azam, U. Naeem, and A. Prugel-Bennett, "Novel centroid selection approaches for KMeans-clustering based recommender systems," *Information sciences*, Vol. 320, pp. 156-189, 2015.

[75] A. Hernando, J. Bobadilla, and F. Ortega, "A non-negative matrix factorization for collaborative filtering recommender systems based on a Bayesian probabilistic model," *Knowledge-Based Systems*, Vol. 97, pp. 188-202, 2016.

[76] E. Q. Da Silva, C. G. Camilo-Junior, L. M. L. Pascoal, and T. C. Rosa, "An evolutionary approach for combining results of recommender systems techniques based on collaborative filtering," *Expert Systems with Applications*, Vol. 53, pp. 204-218, 2016.

[77] Y. Ar, and E. Bostanci, "A genetic algorithm solution to the collaborative filtering problem," *Expert Systems with Applications*, Vol. 61, pp. 122-128, 2016.

[78] M. Nilashi, M. D. Esfahani, M. Z. Roudbaraki, T. Ramayah, and O. Ibrahim, "A multi-criteria collaborative filtering recommender system using clustering and regression techniques," *Journal of Soft Computing and Decision Support Systems*, Vol. 3, No. 5, pp. 24-30, 2016.

[79] C. L. Liao, and S. J. Lee, "A clustering-based approach to improving the efficiency of collaborative filtering recommendation," *Electronic Commerce Research and Applications*, Vol. 18, pp. 1-9, 2016.

[80] N. R. Kermany, and S. H. Alizadeh, "A hybrid multi-criteria recommender system using ontology and neuro-fuzzy techniques," *Electronic Commerce Research and Applications*, Vol. 21, pp. 50-64, 2017.

[81] T. Ebesu, and Y. Fang, "Neural semantic personalized ranking for item cold-start recommendation," *Information Retrieval Journal*, Vol. 20, No. 2, pp. 109-131, 2017.

[82] H. Koohi, and K. Kiani, "A new method to find neighbor users that improves the performance of collaborative filtering," *Expert Systems with Applications*, Vol. 83, pp. 30-39, 2017.

[83] S. Park, and D. Y. Kim, "Assessing language discrepancies between travelers and online travel recommendation systems: Application of the Jaccard distance score to web data mining," *Technological Forecasting and Social Change*, Vol. 123, pp. 381-388, 2017.

[84] M. K. Najafabadi, M. N. R. Mahrin, S. Chuprat, and H. M. Sarkan, "Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data," *Computers in Human Behavior*, Vol. 67, pp. 113-128, 2017.

[85] P. Jomsri, "FUCL mining technique for book recommender system in library service," *Procedia Manufacturing*, Vol. 22, pp. 550-557, 2018.

[86] C. Li, Z. Wang, S. Cao, and L. He, "WLRRS: A new recommendation system based on weighted linear regression models," *Computers & Electrical Engineering*, Vol. 66, pp. 40-47, 2018.

[87] J. Chen, K. Li, H. Rong, K. Bilal, N. Yang, and K. Li, "A disease diagnosis and treatment recommendation system based on big data mining and cloud computing," *Information Sciences*, Vol. 435, pp. 124-149, 2018.

[88] A. S. Tewari, and A. G. Barman, “Sequencing of items in personalized recommendations using multiple recommendation techniques,” *Expert Systems with Applications*, Vol. 97, pp. 70-82, 2018.

[89] I. Viktoratos, A. Tsadiras, and N. Bassiliades, “Combining community-based knowledge with association rule mining to alleviate the cold start problem in context-aware recommender systems,” *Expert systems with applications*, Vol. 101, pp. 78-90, 2018.

[90] R. Ahuja, A. Solanki, and A. Nayyar, “Movie recommender system using K-Means clustering and K-Nearest Neighbor,” In *9th International Conference on Cloud Computing, Data Science and Engineering (Confluence),* IEEE, 2019, pp. 263-268.

[91] B. Hassanpour, N. Abdolvand, and S. Rajaee Harandi, “Improving Accuracy of Recommender Systems using Social Network Information and Longitudinal Data,” *Journal of AI and Data Mining*, Vol. 8, No. 3, pp. 379-389, 2020.

[92] R. C. Tryon, Cluster analysis: correlation profile and orthometric (factor) analysis for the isolation of unities in mind and personality, Edwards’s brother, Incorporated. *Ann Arbor*, 1939.

[93] K. J. Kim, and H. Ahn, “A recommender system using GA K-means clustering in an online shopping market,” *Expert systems with applications*, Vol. 34, No. 2, pp. 1200-1209, 2008.

[94] X. S. Yang, “Firefly algorithms for multimodal optimization,” In *International symposium on stochastic algorithms* Springer, Berlin, Heidelberg, pp. 169-178, 2009.

[95] S. I. Sulaiman, Z. Othman, I. Musirin, and N. S. M. Z. Abidin, “Optimization of an Artificial Neural Network using Firefly Algorithm for modeling AC power from a photovoltaic system,” In *SAI Intelligent Systems Conference (IntelliSys)* IEEE, 2015, pp. 591-594.

[96] R. C. Neath, and M. S. Johnson, “Discrimination and classification,” 2010.

[97] I. Portugal, P. Alencar, and D. Cowan, “The use of machine learning algorithms in recommender systems: A systematic review,” *Expert Systems with Applications*, Vol. 97, pp. 205-227, 2018.

[98] S. L. Lauritzen, Graphical models, Oxford University Press, 1996.

[99] J. S. Breese, D. Heckerman, and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering,” *arXiv preprint arXiv:1301.7363*., 2013.

[100] H. Zhong, H. Zhang, and F. Jia, “Analysis and improvement of evaluation indexes for clustering results,” *EAI Endorsed Transactions on Collaborative Computing*, Vol. 4, No. 13, 2020.

[101] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, “An improved method to construct basic probability assignment based on the confusion matrix for classification problem,” *Information Sciences*, Vol. 340, pp. 250-261, 2016.

[102] the GroupLens Research Project at the University of Minnesota. September 19th, 1997- April 22nd, 1998. Available:https://www.kaggle.com/vedapragnareddy/movie-lens.

[103] C. F. Tsai, and C. Hung, “Cluster ensembles in collaborative filtering recommendation,” *Applied Soft Computing*, Vol. 12, No. 4, pp. 1417-1425, 2012.

[104] S. Renaud-Deputter, T. Xiong, and S. Wang, “Combining collaborative filtering and clustering for implicit recommender system,” In *27th International Conference on Advanced Information Networking and Applications (AINA),* IEEE, 2013, pp. 748-755.

# ارائه یک سیستم توصیه‌گر با استفاده از طبقه‌بندی و خوشه‌بندی با کمک الگوریتم کرم شب‌تاب

**حمیدرضا کوشا[۱،*]، زهرا قربانی[۲] و روشنک نیک‌فطرت[۲]**

[۱]**گروه مهندسی صنایع، دانشگاه فردوسی مشهد، مشهد، ایران.**

[۲] **گروه مهندسی صنایع، دانشگاه صنعتی سجاد، مشهد، ایران.**

ارسال ۲۰۲۱/۰۴/۳۰؛ بازنگری ۲۰۲۱/۰۷/۱۷؛ پذیرش ۲۰۲۱/۱۲/۱۳

**چکیده:**

در دهه‌ی اخیر، تجارت الکترونیکی نقش مهمی در رویکرد مشتریان نسبت به خرید محصولات مختلف داشته است و همچنین باعث رفاه فروشندگان و مزایای بسیاری برای اقتصاد شده است. بدین منظور، به کارگیری سیستمی که با ارائه پیشنهادات متناسب با خواسته‌های یک کاربر (پیشنهادات شخصی سازی شده) او را از مرور تمام آیتم‌ها باز می‌دارد، ضروری می‌باشد. این تحقیق یک سیستم توصیه‌گر دو مرحله‌ای جدید را براساس داده‌های جمعیت شناختی و رتبه‌بندی کاربران در مجموعه داده‌های عمومی MovieLens پیشنهاد می‌کند. در مرحله اول، خوشه‌بندی مشتریان با استفاده از الگوریتم هیبریدی کرم شب‌تاب و الگوریتم K-means انجام شده است. با توجه به توانایی الگوریتم کرم شب‌تاب برای جلوگیری از به دام افتادن در بهینه محلی، که مشکل اصلی K-means است، ترکیب این دو تکنیک منجر به عملکرد بسیار بهتری می‌شود. در مرحله دوم، برای هر خوشه، دو سیستم توصیه‌گر بر مبنای الگوریتم k نزدیکترین همسایگی و طبقه‌بند بیز ساده ارائه شده است. نتایج به‌دست‌آمده با استفاده از معیارهای شاخص Davies-Bouldin، دقت، صحت، پوشش و معیار F مورد ارزیابی قرار گرفته‌اند. نتایج پژوهش حاضر نشان‌دهنده اثربخشی الگوریتم‌های کرم شب‌تاب، K-Means، و K نزدیکترین همسایگی در مقایسه با سایر مدل‌های موجود است.

**کلمات کلیدی**: سیستم توصیه‌گر، الگوریتم کرم شب‌تاب، الگوریتم K-Means، الگوریتم K نزدیکترین همسایگی، الگوریتم بیز ساده، طبقه‌بندی، خوشه‌بندی.