Asian Journal of Distance Education

A Learning Material-Based Recommendation System For E-Learning

İhsan Güneş, Esra Pınar Uça Güneş, Sinan Aydın

Abstract: There may be many kinds of e-learning material presented to learners. However, learners may not have the consciousness or not want to spend time to examine to select the most appropriate one. Recommendation systems seem to be a feasible solution for an efficient learning process both for learners and the service provider. In this study, we propose a learning material-based e-learning recommendation system that considers the learners' learning material preferences and uses the collaborative filtering method for the recommendation system. To obtain realistic results, actual data gained from Anadolum eCampus, the learning management system of Anadolu University Open Education System, were used. In addition, this study aimed to select the most successful algorithm by applying three Collaborative Filtering (CF) algorithms (kNN, k-means and SVD-based CF) in the experiments to keep the efficiency high. As a result, k-means and SVD-based CF algorithms were more successful than kNN-based CF algorithms. In addition, the SVD-based CF algorithm was the most successful regarding speed performance. In conclusion, this system can be used in e-learning settings to recommend learning materials to learners according to their preferences.

Keywords: recommender system, e-learning, collaborative filtering, learning material, data mining

Highlights

What is already known about this topic:

- Recommender systems can be an effective approach for distance learners to improve their learning.
- Recommendation systems are gaining popularity parallel to/with the increasing popularity of personalization and adaptive learning systems in distance education.

What this paper contributes:

- This paper presents the application examples of recommender systems for learning purposes.
- In this article, different models of the Collaborative Filtering (CF) algorithm, widely used in recommendation systems, are applied.
- This study provides a successful method for applying a recommendation system that predicts the learners' material preferences.

Implications for theory, practice and/or policy:

- In addition to material, different suggestions, such as course, learning path and peer can be presented to the learner.
- Different recommendation systems, such as content-based and hybrid, should be applied using different types of data.
- To make better recommendation, many data, such as learner behavior, preferences, course materials activities, assessment results, should be regularly collected in databases.

Published by Asian Society for Open and Distance Education (ASODE), Japan
ISSN 1347-9008 http://www.asianjde.com/

Introduction

E-learning is a method preferred by institutions and learners in education and learning processes due to the many advantages it provides. The increasing popularity and rapid growth of e-learning have changed traditional teaching-learning behavior for teachers and learners. In addition to flexibility and economic benefits, individualization and personalization have gained importance recently as learners may have different abilities, skills, learning styles or backgrounds in different expertise areas. Accordingly, the variety and number of online learning materials offered to learners is increasing. However, learners have difficulty making decisions about choosing good material suitable for their situation. In addition, it becomes more challenging for educators to guide learners in choosing appropriate learning materials. At this point, Zaiane and Luo (2001) state that an e-learning system that can automatically direct learner activities and intelligently produce and recommend learning materials to improve the learning process can be very useful for learners. To better guide learners' learning process, these systems can offer an automated way of getting feedback (Lu, 2004) which is one of the advantages. Based on all these mentioned, it can be said that recommendation systems in e-learning environments can contribute to both educators and learners to use time efficiently during the learning process. In this article, we developed a material-based e-learning recommendation system that considers learner material preferences and uses a collaborative filtering algorithm for the recommendation system.

The Collaborative Filtering (CF) algorithm is the most commonly used approach in recommendation systems. Consumer recommendation, personalized item information, community opinion evaluation, and community criticism are the fundamental features of CF systems. These systems are modeled to enable users to easily locate the things they desire and accommodate the system's excessive information capacity. They calculate the similarity between millions of data sets, using data mining techniques. The CF process includes three major steps. They are computation of similarity, development of neighborhoods, and assessment of recommendations. In a conventional CF process, a user-item matrix is provided and a similarity metric is used to evaluate similarities between each user and active user ‘a’. Then, the most similar user k is selected as the neighbor of a. Finally, a proposal for a target item (q) is presented using a CF algorithm and neighboring data. The prediction (paq) is sent back to active user a.

The general contributions of this article are given below:

- First, data were retrieved from the LMS system on learner ratings.
- The data set was processed and made acceptable for the CF algorithms.
- CF algorithms based on kNN, k-means, and SVD were used. The accuracy of the algorithms' predictions is illustrated graphically.

The rest of the paper is structured as follows: an overview of related research, context and theoretical framework; explanation of kNN, k-means, and SVD-based CF algorithms; the findings and discussion; the conclusion and further work.

Literature

Lu (2004) presented the framework for a personalized learning recommender system. The framework, which may be applied to online teaching and learning sites, is based on two technologies that are interrelated. A multi-attribute evaluation method was utilized to justify a 'learners need, and a fuzzy matching method was used to locate learning resources that best meet the needs of each learner. The author asserted that the use of this approach can enhance the online learning of learners and facilitate the online instruction of big classes with learners from diverse backgrounds. Salehi and Kmalabadi (2012) suggested a new method that considers the characteristics of materials to enhance the quality of recommendations. Therefore, they modeled each learner using a matrix that can account for several material attributes. Additionally, the recommender is adaptive to both individual preferences and fluctuating interests. The significance of factors, such as subject, education level and author, was assessed based on ratings of the resources accessed by the learner. Recommendations are generated through the use of content-based filtering, collaborative filtering, and hybrid methods. Based on the results of studies using real-world datasets, they concluded that the attribute-based strategy satisfies the learner's actual learning preferences in accordance with real-time updated contextual data. Benhamdi et al. (2017) presented NPR eL, a collaborative and content-based filtering-based recommendation method (New multi-Personalized Recommender for e Learning). It was incorporated into a learning environment to give individualized learning content based on the preferences, interests, prior knowledge, and memory capacity of the learners. They utilized the CSHTR (Cold Start Hybrid Taxonomy Recommender) method to circumvent the cold-start issue that plagues recommender systems. Their approach improved CSHTR by incorporating information about the learner's prior knowledge and memory capacity. This makes it possible to tackle the challenging cold-start scenario in which a new learner has no time to read and evaluate documents extraneous to his course.

Liu and Shih (2008) focused on several issues associated with the distribution of instructional materials. According to the authors, these issues are the difficulty of sharing learning resources, the high redundancy of learning materials, and the lack of a course outline. To address these issues, the authors sought to develop an automatic inquiry system for learning materials that takes advantage of the data-sharing and quick searching capabilities of the Lightweight Directory Access Protocol (LDAP) and JAVA Architecture for XML Binding (JAXB). The method is said to be utilized efficiently in education for information collection, processing, digestion, and analysis. Systematic learning, constructive learning, task-based learning, and individualized learning are all facilitated. Salehi et al. (2014) noted that in the existing recommendation algorithms, dynamic interests and multi-preferences of learners as well as multidimensional-attributes of resources are not taken into account concurrently. According to the authors, these algorithms are incapable of utilizing the learner's previous sequential patterns of material access. To address these issues and improve the quality of recommendations, in this study, the authors presented a new framework for a material recommender system based on sequential pattern mining and multidimensional attribute-basedCF.

Using incremental association rule mining, Nadimi-Shahraki (2011) first presented a basic architecture for developing effective personalized learning recommender systems. The author then developed a new way for incrementally mining common patterns from log files, the most computationally intensive step in the proposed architecture. In a single database scan, the content of the log file is recorded using a well-organized tree. The tree can be updated progressively as log files are modified. The author concludes, based on the results of trials, that employing the proposed strategy improves the performance of tailored e-learning material recommenders. Bourkoukou et al. (2016) suggested a personalized e-learning system that considers the personality of the learner and use the CF approach for the recommender system. This model presents modules for personality recognition and learning scenario selection based on the learner's personality. To determine the accuracy of their suggested recommender model's predictions, they utilized an actual E-learning environment. They demonstrate that the proposed method could increase the accuracy of forecasts.

Mawane et al. (2018) proposed clustering based CF system for e-learning framework. According to authors this method improved the classical recommendation system, which is based on ranking similarity and provided effective learning experience for both learners with a problem and those who are struggling to progress. The authors begin by creating clusters based on three learner characteristics: personal information, results, and system involvement. Second, they chose previously unearthed learning objects from the same learning clusters. Using the two sorting criteria, they were provided with a list of the most pertinent learning materials. Mawane et al. (2020) developed a two-step method using unsupervised Kohonen card deep learning to find instrumental approximation of learning styles and deep auto encoder to improve learning resource recommendation. According to the authors, substantial mining is required for this method. Learners had course material, assessment, and learner platform interaction elements. The first phase employed the learner attributes vector, and the learner-content ratings vector determined the best learning resource to recommend. Zriaa and Amali (2021) proposed a comparison analysis of k-NN and k-means clustering to determine the best effective algorithm for prediction in an e-learning recommender system.

Context and Study Area

Turkey's Anadolu University is among the world's largest mega-universities. It is a dual-mode institution with more than one million distance learners. In the 1982-1983 academic year, the Open and Distance Education System of Anadolu University began serving with the Faculty of Open Education as the first faculty in Turkey to offer open and distance education. In 1993, when the Open Education System was restructured, the Faculty of Business Administration and the Faculty of Economics were founded.

Textbooks have been the main learning materials which are designed suitable for self-learning. At the same time, there is always a variety of supporting learning materials and applications depending on the developing technology, changing learning perspectives and the demand of learners. In 1990s, drill and practice software were designed and computer based academic advising centers were opened in some cities. CD-ROMs including training software, video programs, course book (pdf), access to trial exams, toolbox and study guide were sent to learners (Mutlu et al., 2014). From 2000, supporting learning materials have been begun to be served using various ways, such as the internet; online trial exams, virtual classroom application, e-practice, e-book, e-television (transformed tv programs to digitized version), audio book. Then, different e-learning platforms were used in time to present these materials and some other services, such as e-support systems. First one was the e-Learning Portal that is enriched with innovative e-learning applications focused on the individual learning needs of learners. In the following years, new formats for learning materials were used, such as e-course, e-tutoring, interactive e-books, e-seminars. With the increasing number of content types and numbers, Beta e-Learning Portal was launched in 2015-2016 to make it easier for learners to access content (Kumtepe et al., 2017). The Beta e-Learning Portal had a user-friendly interface that provided easy access to the course content from a single interface. The learning materials included e-book, summary chapter, paper test, essay exam, past exam questions, exercise questions, trial exam, audio book, audio summary, interactive e-book, interactive e-lecture, e-seminar, animation videos, TV videos, summary chapter video, 1 question 1 answer videos.

Lastly, Anadolum eCampus is being used by the learners in the Open Education System from spring 2015-2016 academic year. It is a learning management system that offers open and distance learning services as a whole package and has a modular structure. The foremost factor to use this system is to increase learners' motivation and interaction in the learning process. The system consists learning management system, learning analytics, live course platform and mobile application.

This digital campus system provides access to learning materials, lectures, announcements, calendars, content prepared by RadioA and online learning communities. The system is intended to enrich learning experiences. Considering learners may have different features and learning preferences, learning materials are presented in different genres on Anadolum eCampus. Twenty seven different learning materials and services/applications presented at Anadolum eCampus can be classified as text (pdf) type, audio type, video type and interactive type. Learning materials/services are listed in Table 1:

Table 1. Learning Materials/Services Presented in Anadolum eCampus System

Text (pdf) type	**Audio type**	**Video type**	**Interactive type**
Course book	Audio Book	Course Description Video	Question and Play Application
Unit Text	Audio Summary	eSeminar	Interactive eBook
Unit Summary		Unit Lecture Video	Interactive eLecture
Learn using Questions		Unit Lecture Video (TV)	Trial Exams
Solved Questions		1 question 1 answer	Exercises
Paper Test		Animation Videos	Discussion Forums
Trial Exam		Interactive Video	eCanteen
Past Exam Questions		TRT School Videos	Online Learner Communities
			Volunteer Quality Ambassadors

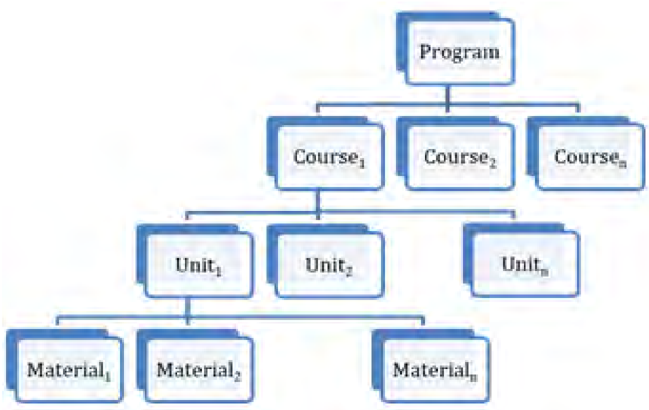
The system provides a rich content for learners about learning materials and services and analyzes the interaction of the learners with these via logs, questionnaires, interviews and rates. As a result, alternatives are assessed to determine what kinds of materials and services will/should be available to learners for future action plans. However, it is not yet possible to make individual recommendations for these services which are produced and presented to the masses. The courses in the Anadolum eCampus System are designed unit-based. Most of the courses consist of eight units, although there are a few courses including 6, 7 or 9 units. Figure 1 shows the course-unit-material structure.

According to Klašnja-Milićević et al. (2017), recommender systems must aggregate personal preference information, such as the user's purchase history, clickstream data, and so on, to provide personalized recommendations tailored to the user's individual needs. Users' product preferences are traditionally referred to as ratings, and the authors distinguish between two distinct forms of ratings:

1. Explicit ratings that ask users to declare their preferences for a specific item, typically using 5- or 7-point Likert scales.
2. Implicit ratings via collecting preference behavior from observations of user behavior.

In the Anadolum eCampus system, learners can share their satisfaction using between 1 and 5 points to evaluate the material. In 2016-2017 academic year, this rate system was only available for the materials below. There is also a field where they can express their views about the material. In this study, these actual rate values were used. The rate values given by 6182 users to 9 different materials were used.

Figure 1. Course-unit-material structure



- eSeminar Records: Learning materials based on the interaction of field specialists and learners simultaneously through the virtual classroom. The eSeminar records are also uploaded to the Anadolum eCampus for watching after the course.
- Trial Exam (Midterm-PDF): Test exams prepared composed of questions about the units that the learners are responsible for examinations and the answer key. They are course-based and covers midterm exam responsibilities.
- Trial Exam (End of Term-PDF): Test exams prepared composed of questions about the units that the learners are responsible for examinations and the answer key. They are course-based and covers all of the units.
- Course book (PDF): Learning materials prepared by subject experts for each course and presented to learners as the main source.
- Learn using Questions: A unit-based material that the important points of the relevant unit are prepared as open-ended questions and presented together with their answers.
- Unit Lecture Video: Videos composed of narratives summarizing the unit prepared by field experts.
- Unit Lecture Video (TV): Unit based narrative videos, of which scenarios are prepared by the field experts and which are designed and edited by TV Production Center.
- Paper Test: Unit-based tests consisting multiple-choice questions and answer keys for the unit.
- Unit Summary: Materials that covers the important points and enable learners to prepare for the units

We have nearly 1200 courses in Anadolum eCampus. In this platform, we obtain rates for nine different materials. This study aims to identify the materials that learners may prefer. Accordingly, if the target item evaluation prediction value for the active user is calculated as high as 4 or 5, the system will recommend this material to this learner.

Theoretical Framework

There exists a diverse array of recommendation algorithms and methodologies, which can generally be classified into three overarching categories: collaborative filtering, content filtering, and hybrid filtering. CF is a commonly employed technique in recommender systems for the purpose of estimating user preferences. The term "CF" was initially coined by the Tapestry system (Goldberg et al., 1992), which was originally conceived for the purpose of email filtering during the early 1990s. CF is a technique that relies on the aggregation of user preferences and uses a weighted average methodology to assess predictions based on the ratings of comparable items (Herlocker et al., 2004).

Different algorithms can be used in recommendation systems to increase the performance and accuracy of the CF algorithm. The data in the recommendation system can be modeled with these algorithms (e.g., Classification, Clustering, and Matrix factorization) so that the system can work faster and more accurately. In this study, kNN, k-means, and SVD-based CF recommendation systems, which are convenient for our data and also easier to implement, were used.

CF algorithms

Basically, there are three groups of CF algorithms: memory-based, model-based and hybrid. Memory-based CF methods make predictions using all or part of the user database. The following operations are performed by the neighborhood-based CF algorithm, also known as the extended memory-based CF algorithm: It calculates the similarity or weight, which reflects the distance, correlation or weight between any two entities (users or items), creates a neighborhood, and generates a prediction for the active user by taking a weighted average or a simple weighted average of all ratings for a given item or user

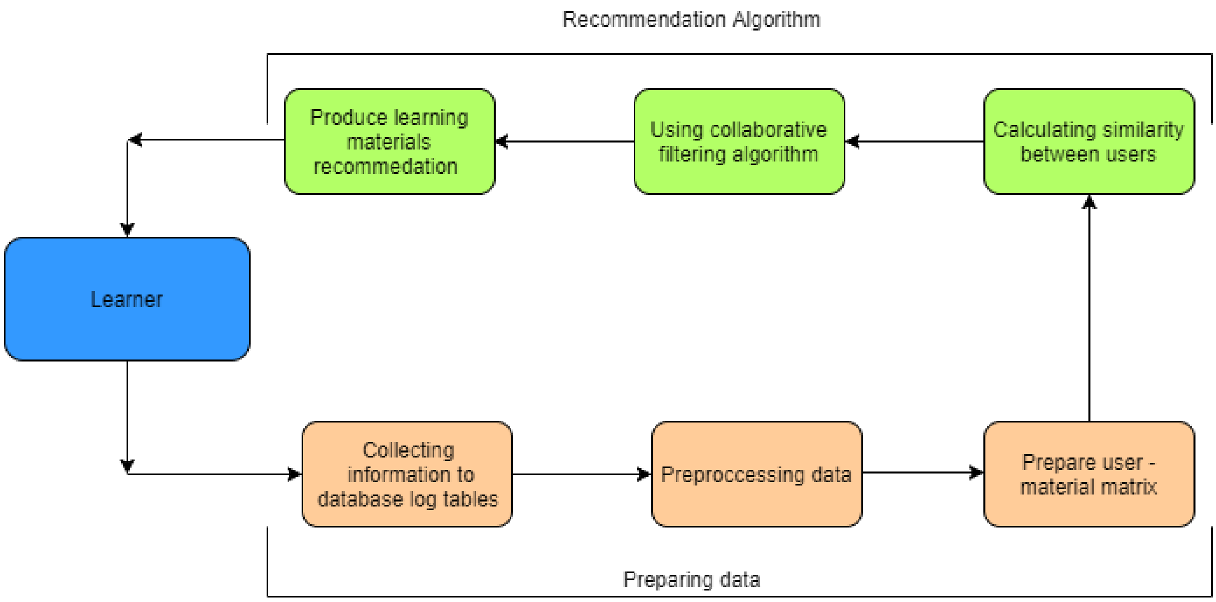
(Herlocker et al., 2004). Once similarities are calculated, comparable people or similar objects (nearest neighbors) need to be found when the goal is to provide the best N recommendations. Neighbors aggregate the top-selling N-items as recommendations (Su and Khoshgoftaar, 2009).

Model-based CF algorithms use system data (user ratings) to construct a model, which is then employed to generate recommendations. Existing model-based algorithms include cluster models, probabilistic models, Bayesian networks, rule-based systems, and techniques for dimensionality reduction. Clustering, for example, is the initial endeavor to divide a dataset into user groups. The bisection k-means algorithm (Steinbach et al., 2000) is a variant of the k-means clustering technique used for clustering. A Bayesian network model expresses a probability model for the CF problem (Breese et al., 1998). Billsus and Pazzani (1998) detail the limitations of CF approaches that employ a learning algorithm. The proposed techniques utilize Singular Value Decomposition (SVD) to reduce the dimensions of the initial matrix of user ratings. To enhance the functionality of the CF algorithm, SVD is used to reduce the dimension (Sarwar et al., 2000). Russell and Yoon (2008) propose utilizing various wavelet transformations (DWT) for the proposed systems. Before a time estimation, the data are transformed into these systems and considerably condensed to make them sufficiently long. A novel method based on incremental SVD and the generalized Hebbian algorithm is proposed (Polezhaeva, 2011). When a new user or new object is introduced, the new algorithm efficiently revises user/item profiles. It is unnecessary to store the initial data matrix.

To resolve the limitations of pure CF, hybrid methods combine the advantages of memory and model-based approaches. The recommendation function of these algorithms is typically preferable to that of memory-based or model-based CF methods. Probabilistic memory-based CF (Yu et al., 2004) combines memory and model-based techniques. This method predicts the poster using a hybrid model derived from a compilation of registered user profiles and user ratings. Pennock et al. (2000) describe personality diagnosis as a descriptive hybrid CF approach that combines memory and model-based CF algorithms and retains some of the advantages of both.

In this study, three different algorithms were applied and their results were compared. These were k-Nearest Neighbors (kNN), k-means and Singular Value Decomposition (SVD) based CF algorithms. Figure 2 shows our CF recommendation system model process flows.

Figure 2. CF recommendation system model



CF recommendation system model can be summarized in three steps:

- Data collection and preprocessing data: Learner rate information is collected using LMS. This data is collected in the log database. This rate data from the log database is prepared on a learner basis. The rate that each learner gives to each item is extracted. As a result, the data are organized and a user-material matrix is obtained.
- Calculating similarity between users: This is performed for each user. Each user is considered and similarity to other users is calculated. The Pearson Correlation formula is used for these similarity calculations. The similarity calculations are explained in more detail below.
- Computing prediction and recommendation: The final step is to estimate the rate which the user will give for any material. In the following section, the prediction formula is explained in detail. If the estimate is as high as 4 or 5, this material may be presented as a recommendation to the user. On the contrary, if it is low, it is not offered as a suggestion.

Methodology

In this study, learner course material evaluation scores in the Open Education System were used as a data set. Experiments were carried out using the data in the three algorithms described below according to the CF recommendation system model (Figure 2) mentioned in the theoretical framework.
One memory-based and two model-based algorithms were used in this article. The kNN based CF algorithm we use for this article is the most widely used memory-based CF algorithm. SVD and k-means based CF algorithms are model-based CF algorithms.

In this study, user data were collected through the Open Education LMS system. The evaluation scores given by 6812 users to 9 different course materials were used. Learners were able to give evaluation scores between 1-5 for the materials in their courses. Course materials with the most used and having the highest number of evaluation scores were selected. It is aimed to estimate the evaluation score given by each user to the materials using three different algorithms and determine which algorithm is more successful. To measure the success of the algorithms, each user is taken as an active user, respectively, and the course materials for which the evaluation score is given are selected as the target item. The CF algorithm recalculates the evaluation score for the active user using the most similar users' data who have rated the selected course material. This process is performed for all users in each algorithm. Accordingly, the difference (MAE) between the actual value given to the materials by each user and the estimated value is calculated. Afterwards, the most successful algorithm can be selected and course material can be recommended to users who are new to the system according to this model. The explanations regarding the algorithms and MAE are given below.

kNN-based CF

Votes are stored as a user-product matrix in a typical CF system. The U_{nxm} matrix contains the product preferences of *n* users for *m* products. An active user (a) who wishes to provide a prediction for the target product *q* transmits his or her current ratings to the system during an online contact with a CF system. The CF prediction method is a two-step procedure that (1) identifies adjacencies by computing similarities between *a* and all other system users and (2) approximates a weighted average based on the weights of adjacent users on *q*. Several methods are used to determine similarities between *a* user and any other user. PCC (Pearson Correlation Coefficient) is one of the finest measures of similarity. The GroupLens project (Resnick et al., 1994) was the first to describe this formulation. PCC is defined as the basis for calculating the weight in this investigation. Equation 1 represents how to determine the correlation between active user *a* and user *u* (Breese et al., 1998).

$$w_{au} = \frac{\sum_{j \in M}(v_{aj} - \overline{v_a})\,(v_{uj} - \overline{v_u})}{\sqrt{\sum_{j \in M}(v_{aj} - \overline{v_a})^2}\,\sqrt{\sum_{j \in M}(v_{uj} - \overline{v_u})^2}} \tag{1}$$

where v_{aj} and v_{uj} represent the votes cast by users *a* and *u* for item *j*. $\underline{v_a}$ and $\underline{v_u}$ represent the average votes cast by users *a* and *u*, respectively. *M* represents the number of things rated by both *a* and *u*. After computing similarity, the *k* users most similar are identified as neighbors (Herlocker et al., 2004).

GroupLens introduced an automatic CF system based on a kNN algorithm (Konstan et al., 1997; Resnick et al., 1994). GroupLens provided tailored predictions for news articles on Usenet. The initial GroupLens system uses PCC to measure user similarities. In the first method, the k most comparable neighbors are chosen. Equation 2 generates a prediction for *a* on *q*, denoted by p_{aq}, as a weighted average of the scores of the ones close to each other on *q*.

$$p_{aq} = \overline{v_a} + \frac{\sum_{u=1}^{N}(v_{uq} - \overline{v_u})w_{au}}{\sum_{u=1}^{N} w_{au}} \tag{2}$$

The similarity weight between a and u is denoted by w_{au}.

k-means clustering-based CF

The k-means algorithm is a well-known algorithm for clustering related objects. The beginning object for the cluster center is chosen at random, and each object is subsequently assigned to the nearest cluster based on the similarity measure. As the average of the items, the cluster centers are recalculated at each stage. This procedure concludes when all cluster members remain unchanged. The model-based CF system use k-means clustering to address issues, such as scalability. In offline mode, the k-means-based CF algorithm groups user profiles into k clusters. When the active user a requests the prediction item *q*, the similarity of the server a user to each cluster center is first determined. The Pearson correlation similarity measure is implemented as follows.

$$w_{aC} = \frac{\sum_{j=1}^{m}(v_{aj} - \overline{v_a})(v_{Cj} - \overline{v_C})}{\sigma_a \times \sigma_C \sqrt{\sum_{j \in M}(v_{aj} - \overline{v_a})^2}\,\sqrt{\sum_{j \in M}(v_{Cj} - \overline{v_C})^2}} \tag{3}$$

We identify the cluster that is most comparable to user a. When a cluster is discovered, the similarity between user a and cluster members is computed. The clustering models in CF only reduce similarity by assessing the similarity between user a and clusters, as opposed to all system users. Lastly, we select the largest k user in that cluster, the largest k group that is most similar to user a.

User a's item q prediction is weighted average of neighbors' scores.

$$p_{aq} = \overline{r_a} + \frac{\sum_{u=1}^{N}(v_{uq} - \overline{v_u})w_{au}}{\sum_{u=1}^{N} w_{au}} \tag{4}$$

N is represented number of neighbors chosen corresponding cluster, w_{au} is the similarity between active user *a* and neighbors 'user, v_{uq} is rating score of neighbors' on item *q*.

SVD-based collaborative filtering

When the system has a large number of user-item data, traditional CF algorithms will be subject to severe scalability problems. SVD can be used for CF algorithms to overcome the scalability problem. SVD reduces the size of the database that contains user/item rates and increases the performance of the CF algorithm. SVD-based CF algorithm is used in this study.

SVD is a common matrix factorization method that decomposes a $n \times m$ matrix *M* into three matrices with the formula $M = USV^T$, where *U* and *V* are two orthogonal matrices of size $n{\times}r$ and $m{\times}r$, respectively, and *r* is the rank of the matrix *M*. S is a $r \times r$ diagonal matrix with all singular values of matrix *M* as diagonal elements. The SVD-based CF algorithm is proposed by Sarwar et al. (2000). In the beginning, the sparse user-item matrix *M* is populated with average item values. The filled matrix is normalized (M_{norm}) using the application of z-score, and the M_{norm} matrix is decomposed into three matrices as *U*, *S*, and *V* with the use of SVD. To generate matrix S_k, r × r matrix *S* is reduced by picking only the *k* greatest diagonal values, $k <= r$. $U_k \sqrt{S_k}$ and $\sqrt{S_k}\, V_k^T$ are then calculated. The prediction for user a on item q, the scalar product of the ath row of $U_k\sqrt{S_k}$ and the qth column of $\sqrt{S_k}V_k^T$, is calculated, and the result is denormalized as follows:

$$P_{aq} = \overline{v_a} + \sigma_a[\, U_k\sqrt{S_k}(a).\sqrt{S_k}V_k^T(q)] \tag{5}$$

MAE (Mean Absolute Error)

This study used the MAE metric, which calculates the absolute difference between the actual evaluation score given to the material and the predicted value. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are often used as statistical accuracy metrics. Mean Absolute Error (MAE) is a frequently employed metric for assessing recommendation systems. It measures the average absolute difference between predicted and actual scores or ratings provided by users. A lower MAE value indicates that a recommendation system is more accurate. MAE is the most popular and widely used method; a measure of deviations from the user's particular value of recommendation. It is computed as follows:

$$\mathrm{MAE} = \frac{1}{N}\left(\sum_{u}^{i} |p_{ui} - r_{ui}|\right) \tag{6}$$

where P_{ui} represents the predicted rating for user u on item *i*, r_{ui} represents the actual rating, and *N* represents the total number of ratings for the item collection. The MAE indicates the accuracy with which the recommendation engine predicts user ratings.

Limitations

When we examine the literature, it is very rare to find real data sets that can be studied in this field. Since we have a large distance education system, we are able to work with real data in this study. However, the recommendation system we have developed may not be able to find course materials to recommend due to the lack of sufficient number of course materials and the insufficient number of evaluations of some materials. To avoid this situation, the variety of course materials should be increased and learners' interest in course materials should be increased. In addition, only learner material evaluation data were used in this study. In addition, other data, such as learners' behavior in the LMS system, and learner profile information (e.g., age, sex, and location), can be examined and suggestions can be offered to learners.

Findings and Discussions

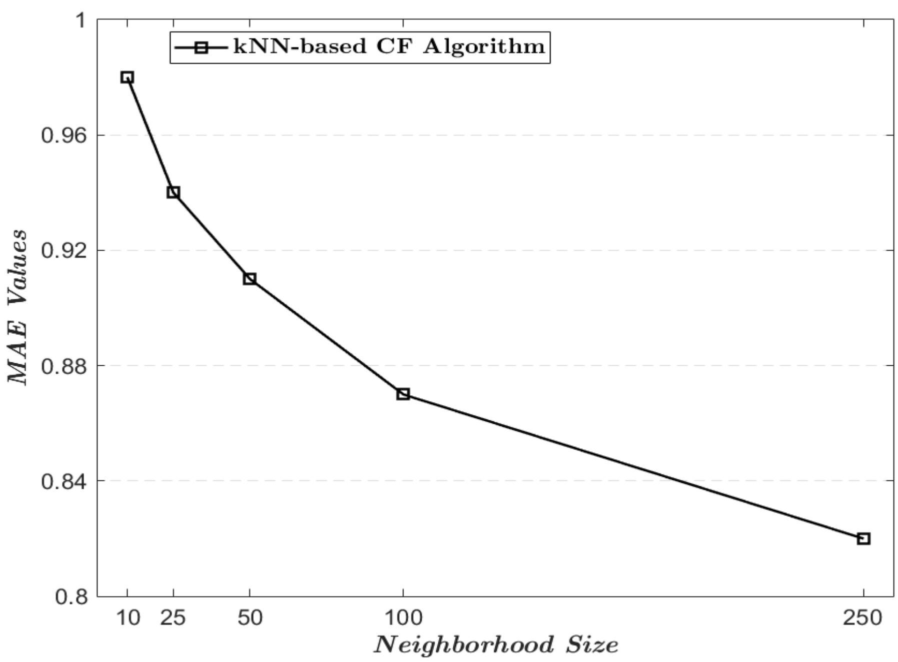
The data collected for our experiments consisted of user/item ratings provided for the e-learning materials. Our experimental set consisted of 6812 users who rated for nine different materials. There were 15,961 votes, which had been submitted by 6,812 users. The data within the matrix, specifically the calculation of 6812 multiplied by 9 resulting in 61.308, indicated that the matrix was approximately 20% full. The ratings within the matrix were discrete, ranging from 1 to 5.

The experiments were conducted utilizing the all-but-one technique. In each iteration, one user was designated as the active user, while the remaining users form the training group. Furthermore, a set of target items was generated, comprising nine distinct material types. Mean Absolute Error (MAE) values were computed for each user and item in the dataset.

Three CF algorithms were applied in the experiments. As mentioned above, these were kNN, k-means and SVD-based CF algorithms. For these three methods, the MAE values were calculated with the same data set. these three methods had different parameters that affected performance. These were neighbors in kNN, cluster size in k-means and matrix reduction size in SVD.

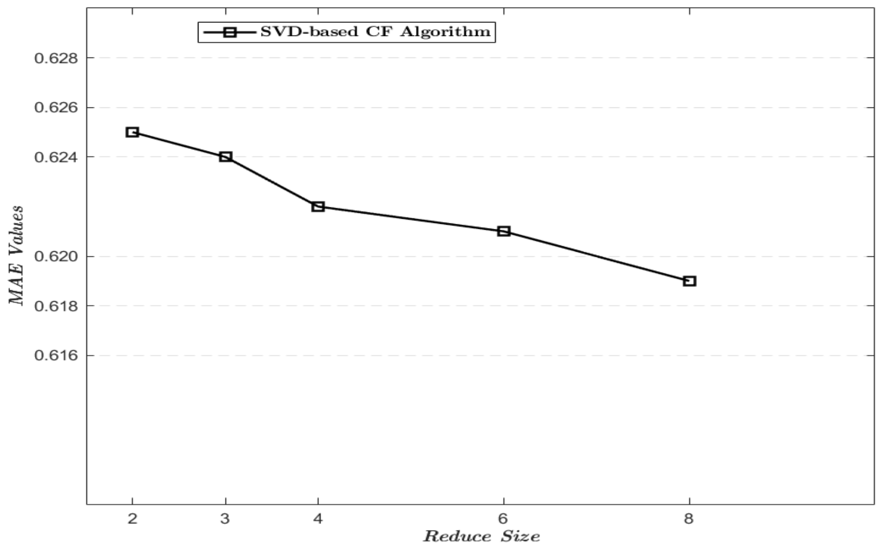
When we examined the kNN algorithm, it was seen that while the number of neighbors increased, the value of MAE decreased (Figure 3). The best result was obtained with a value of 0.82 at a neighbor number of 250.

Figure 3. MAE Values for varying number of neighborhoods



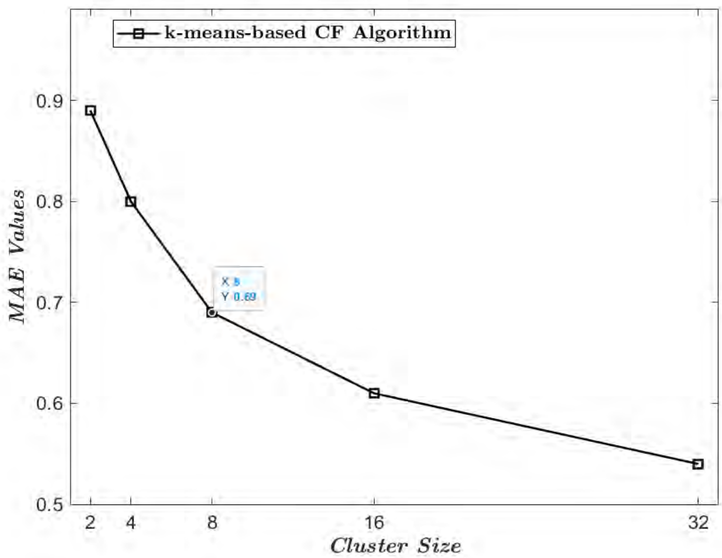
For the SVD-based CF algorithm, the MAE change was analyzed using the matrix download parameter. As can be seen, the MAE values were calculated by reducing to matrix size 2, 4, 6, 8. reduced *d* dimensional space. As can be seen in the Figure 4, the best result will be obtained in 8-dimensional matrix size. In this case, it can be explained that the MAE value is good because the loss of data is small in the 8 dimensional matrix.

Figure 4. MAE Values for varying reduce size



In the k-means-based CF algorithm, the effect of the variation of the cluster size on the MAE was analyzed. As the cluster size increased, the MAE value decreased (Figure 5). The best MAE value was obtained with a cluster size of 32. This can happen because in the smaller clusters the most similar elements fall into the same cluster. Only the neighbors n this set are taken into account when calculating the prediction.

Figure 5. MAE Values for varying number of cluster



As a result, when all methods were compared, it was seen that k-means and SVD-based CF algorithms were more successful than kNN-based CF algorithms. In addition, the SVD-based CF algorithm was the most successful regarding speed performance. This is due to the matrix size reduction of SVD. As a result, the MAE values range from 0.6 to 1.0.

The application in this study can estimate the rate value to be given by the learner to material with an error margin as much as the MAE value. The lower the MAE value is, the better the prediction quality is. If, after the calculated estimates, the values for rates are high, such as 4 and 5, this material can be presented to the learner as a recommendation. However, if the values are low, such as 1 and 2, they will not be presented as recommendations.

In general, SVD and k-means CF algorithms were close to each other in terms of prediction quality. However, it was also observed that these two algorithms obtained much better results than kNN-based CF. In this study, course material rate data was used. When we examined other studies using different data, it was observed that the SVD and k-means CF algorithms were generally more successful than the kNN-CF algorithm.

In their study, Ahmed and Letta (2023) studied book recommendation as a university library service using Model-based SVD and kNN CF algorithms. Consistent with the findings in our study, they reported that Matrix factorization (SVD) outperforms kNN models in generating CF proposers. The accuracy score of the proposed SVD-based model was 85%, which was better than the kNN model (53%). In another study, Ba et al. (2013) proposed a new CF algorithm that combined SVD algorithm and clustering algorithm and compared this new algorithm with classical SVD and kNN. They used the 100k dataset from the movie website MovieLens, which included about 100k ratings of 1682 movies from 943 users. The method developed by the authors had the lowest MAE value, followed by SVD and kNN algorithms, respectively. In addition, Hssina et al. (2021) evaluated the performance of their Hybrid SVD-kNN algorithm, and compared it with other algorithms regarding several metrics. According to the evaluation results, the SVD algorithm outperformed the kNN algorithm and the random predictor. The kNN algorithm was nearly as good as SVD, but SVD had a slightly better performance in all metrics.

However, in a study (e-Learning Course Recommender System) that focused on the MI-based models, such as kNN, SVD and neural network–based collaborative filtering (NCF) models, Jena et al. (2023) found kNN to perform better as compared to other models.

The MAE values measured in studies using similar algorithms may vary depending on data quality, type and size. As a general comment, SVD and k-means-based CF algorithms were more successful than kNN-based CF algorithms, as in this study. Since each algorithm has its own requirements and often does not meet every criterion, the purpose of the application and the amount and type of data to be processed are decisive when choosing an algorithm. When similar studies in the field of distance education are analyzed, it is seen that SVD, k-means and kNN algorithms are suitable for our data.

Conclusion and Suggestions

Issues related to personalization in the learning process have been discussed extensively in the last decades and continue to remain the focus of interest of many researchers to date. In this study, we propose a material-based e-learning recommendation system that takes into account the learner's material preferences and uses the common filtering method for the recommendation system.

In our study, kNN, k-means and SVD-based CF algorithms were applied. MAE values ranging from 0.54 to 0.98 were obtained for all three algorithms. The lower the MAE value, which shows the difference between the predicted value and the actual value, the more successful the algorithm is. As our application results showed, k-means and SVD-based CF models obtained more successful results. SVD and k-means algorithms are more sophisticated and robust than kNN algorithm. As indicated in the Finding and Discussion section above, other studies have reported the efficacy of SVD and k-means algorithms, which aligns with our own findings. As explained above, the k-means CF algorithm is more successful, especially when we increase the number of clusters. The SVD algorithm is the fastest algorithm since it works by reducing the size of the user-item rating matrix. In the future, we plan to implement different algorithms so that we can achieve better results. In addition, new models can be created by combining the advanced features of different algorithms to achieve better results to be applied in further studies.

The actual rate values given to the materials used in our Open Education System are used. The results of the experiments were very successful at predicting the learners' material preferences. In our first study, nine materials were used. However, as we increase the number of materials in our e-learning portal, the materials we will use in the recommendation system will also increase. Besides, we plan to offer different recommendations to learners, such as course, learning path and peer. The use of recommender systems in e-learning is a rapidly developing field. However, the greatest challenge in this area is the inability to find the actual test data. We are more advantageous in data collection than other universities Because we have about 1.2 million learners enrolled in our Open Education System and we have a very advanced LMS infrastructure from which we can obtain data. In our further studies, we aim to analyze all log data belonging to the learners and provide individualized recommendations online.

Since there are not the same types of materials for each course or there are few materials in some courses, there may be limitations in the materials to be recommended to the learner. In addition, the evaluation criteria for each material may be different. For example, in future works different sub-evaluation criteria can be added for textual content and video-based content.

We have developed our recommendations by looking at the general material type preference of learners in this study. That is, we have assumed that a learner gives a rate to materials according to his/her learning style. However, there are many other factors that affect the rate of materials, e.g., the presenter's expression in video-type materials. Therefore, a recommendation system can be developed to propose course-based materials in further works.

References

Ba, Q., Li, X., & Bai, Z. (2013). Clustering collaborative filtering recommendation system based on SVD algorithm. 4th International Conference on Software Engineering and Service Science (pp. 963-967).

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, *46*, 109-132. https://doi.org/10.1016/j.knosys.2013.03.012

Benhamdi, S., Babouri, A., & Chiky, R. (2017). Personalized recommender system for e-Learning environment. *Education and Information Technologies*, *22*(4), 1455-1477. https://doi.org/10.1007/s10639-016-9504-y

Billsus, D., & Pazzani, M. J. (1998). Learning Collaborative Information Filters. International Conference on Machine Learning (ICML'98). https://dl.acm.org/doi/10.5555/645527.657311

Bourkoukou, O., El Bachari, E., & El Adnani, M. (2016). A personalized e-learning based on recommender system. *International Journal of Learning and Teaching*, *2*(2), 99-103. doi: 10.18178/ijlt.2.2.99-103

Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence.* http://research.microsoft.com/pubs/69656/tr-98-12.pdf

Desrosiers, C., & Karypis, G. (2011). A comprehensive survey of neighborhood-based recommendation methods. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P. (Eds) *Recommender systems handbook* (pp. 107-144). Springer, Boston, MA. https://doi.org/10.1007/978-0-387-85820-3_4

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, *35*(12), 61-70. https://doi.org/10.1145/138859.138867

Grcar, M. (2004). User profiling: Collaborative filtering. *Proceedings of SIKDD 2004 at Multiconference IS 2004*, 75. https://ailab.ijs.si/Dunja/SiKDD2004/Papers/MihaGrcar-CollaborativeFiltering.pdf

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS), 22*(1), 5-53. https://doi.org/10.1145/963770.963772

Hill, W., Stead, L., Rosenstein, M., & Furnas, G. (1995). *Recommending and evaluating choices in a virtual community of use*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA. https://doi.org/10.1145/223904.223929

Hssina, B., Grota, A., & Erritali, M. (2021). Recommendation system using the k-nearest neighbors and singular value decomposition algorithms. *International Journal of Electrical and Computer Engineering (IJECE), 11*(6), 5541-5548.

Jena, K. K., Bhoi, S. K., Malik, T. K., Sahoo, K. S., Jhanjhi, N. Z., Bhatia, S., & Amsaad, F. (2023). E-learning course recommender system using collaborative filtering models. *Electronics, 12*(1), 157. https://doi.org/10.3390/electronics12010157

Klašnja-Milićević, A., Vesin, B., Ivanović, M., Budimac, Z., & Jain, L. C. (2017). *Recommender systems in e-learning environments.* In E-Learning Systems (pp. 51-75). Springer, Cham. https://doi.org/10.1007/978-3-319-41163-7_6

Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). GroupLens: applying collaborative filtering to Usenet news. *Communications of the ACM, 40*(3), 77-87. https://doi.org/10.1145/245108.245126

Kumtepe, A.T., Büyük, K., Güneş, İ., Öztürk, A., Tuna, G., Gümüş, S., & Atak, N. (2017). Kitlesel uzaktan eğitimde öğrenen-içerik etkileşimi: Anadolu Üniversitesi Açıköğretim Sistemi örneği. Açıköğretim Uygulamaları ve Araştırmaları Dergisi*,* *3*(2), 9-36. https://dergipark.org.tr/tr/pub/auad/issue/34117/378437

Liu, F.-j., & Shih, B.-j. (2008). E-learning activity-based material recommendation system. *Interactive Technology and Smart Education, 4*(4), 200-207. https://doi.org/10.1108/17415650880001105

Lu, J. (2004). *A personalized e-learning material recommender system*. International Conference on Information Technology and Applications. https://opus.lib.uts.edu.au/bitstream/10453/3057/1/2004001806.pdf

Mawane, J., Naji, A., & Ramdani, M. (2018). Clustering collaborative filtering approach for Diftari E-Learning platform'recommendation system. *Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications* (pp. 1-6). https://doi.org/10.1145/3289402.3289535

Mawane, J., Naji, A., & Ramdani, M. (2020) Unsupervised deep collaborative filtering recommender system for e-learning platforms.*International Conference on Smart Applications and Data Analysis* (pp. 146-161). Springer, Cham. 10.1007/978-3-030-45183-7_11

Mutlu, M.E., Özöğüt Erorta, Ö., Kip Kayabaş, B. & Kayabaş, İ. (2014). Anadolu Üniversitesi Açıköğretim Sisteminde e-öğrenmenin gelişimi, In A.E. Özkul, C.H. Aydın, E.G. Kumtepe, ve E. Toprak, (Eds.), *Açıköğretimle 30 Yıl* (s.1-58). Anadolu Üniversitesi Yayınları.

Nadimi-Shahraki, M. H. (2011). Efficient personalized e-learning material recommender systems based on incremental frequent pattern mining. *The 2011 International Conference on Information and Knowledge Engineering.* https://www.researchgate.net/publication/225090504_Efficient_personalized_e-learning_material_recommender_systems_based_on_incremental_frequent_pattern_mining

Pennock, D. M., Horvitz, E., Lawrence, S., & Giles, C. L. (2000). Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach. *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*. https://clgiles.ist.psu.edu/pubs/UAI-2000-personality-diagnosis.pdf

Polezhaeva, E. (2011). Incremental methods in collaborative filtering for ordinal data. *International Conference on Pattern Recognition and Machine Intelligence.* https://doi.org/10.1007/978-3-642-21786-9_73

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: an open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM conference on Computer supported cooperative work.* https://doi.org/10.1145/192844.192905

Russell, S., & Yoon, V. (2008). Applications of wavelet data reduction in a recommender system. *Expert Systems with Applications, 34*(4), 2316-2325. https://doi.org/10.1016/j.eswa.2007.03.009

Salehi, M., Kmalabadi, I. N., & Ghoushchi, M. B. G. (2014). Personalized recommendation of learning material using sequential pattern mining and attribute based collaborative filtering. *Education and Information Technologies, 19*(4), 713-735.https://doi.org/10.1007/s10639-012-9245-5

Salehi, M., & Kmalabadi, I. N. (2012). A hybrid attribute–based recommender system for e–learning material recommendation. *IERI Procedia, 2*, 565-570. https://doi.org/10.1016/j.ieri.2012.06.135

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). *Application of dimensionality reduction in recommender system - A case study*.Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/215429.

Steinbach, M., Karypis, G., & Kumar, V. (2000). *A comparison of document clustering techniques.* KDD Workshop on Text Mining.

Su, X., & Khoshgoftaar, T. M. (2009). *A survey of collaborative filtering techniques.* Advances in artificial intelligence. https://doi.org/10.1155/2009/421425

Yu, K., Schwaighofer, A., Tresp, V., Xu, X., & Kriegel, H.-P. (2004). Probabilistic memory-based collaborative filtering. *IEEE Transactions on Knowledge and Data Engineering, 16*(1), 56-69. doi: 10.1109/TKDE.2004.1264822

Zaiane, O. R., & Luo, J. (2001). Web usage mining for a better web-based learning environment. *Proceedings of Conference on Advanced Technology for Education* (pp. 60-64).

https://www.researchgate.net/publication/2373061_Web_Usage_Mining_for_a_Better_Web-Based_Learning_Environment

Zriaa, R., & Amali, S. (2021). A comparative study between k-nearest neighbors and k-means clustering techniques of collaborative filtering in e-learning environment. *Proceedings of the Third International Conference on Smart City Applications* (pp. 268-282). Springer, Cham. 10.1007/978-3-030-66840-2_21

About the Authors

- Ihsan Gunes (Corresponding author); igunes@eskisehir.edu.tr; Eskisehir Technical University, Turkey; https://orcid.org/ 0000-0001-5932-8068
- Esra Pınar Uca Gunes; epug@eskisehir.edu.tr; Eskisehir Technical University, Turkey; https://orcid.org/0000-0003-1750-1998
- Sinan Aydın; snaydin@anadolu.edu.tr; Anadolu University, Turkey; https://orcid.org/0000-0003-3014-1384

Author’s Contributions (CRediT)

İhsan Güneş: Conceptualization, Data curation, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing.; Esra Pınar Uça Güneş: Validation, Visualization, Writing – original draft, Writing – review & editing.; Sinan Aydın: Resources, Formal Analysis, Writing – original draft, Writing – review & editing

Acknowledgements

Not applicable

Funding

Not applicable.

Ethics Statement

Approval for the use of the course material evaluation data used in this article was obtained from the Anadolu University Open Education Faculty Dean's office.

Conflict of Interest

The authors do not declare any conflict of interest.

Data Availability Statement

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Suggested citation:
Güneş, İ., Uça Güneş, E. P., & Aydın, S. (2023). A Learning Material-Based Recommendation System For E-Learning. *Asian Journal of Distance Education, 18*(2), 129-145. https://doi.org/10.5281/zenodo.8352369