Expert Systems With Applications 199 (2022) 117092

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Knowledge discovery for course choice decision in Massive Open Online Courses using machine learning approaches

Mehrbakhsh Nilashi [a,b,i,*], Behrouz Minaei-Bidgoli [b], Abdullah Alghamdi [c], Mesfer Alrizq [d], Omar Alghamdi [e], Fatima Khan Nayer [f], Nojood O Aljehane [g], Arash Khosravi [h], Saidatulakmal Mohd [j]

[a] Centre for Global Sustainability Studies (CGSS), Universiti Sains Malaysia, 11800 George Town, Malaysia
[b] School of Computer Engineering, University of Science and Technology, Tehran, Iran
[c] Information Systems Dept, College of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia
[d] Information Systems Department, College of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia
[e] Business Administration Dept, Applied College, Najran University, Najran, Saudi Arabia
[f] Artificial Intelligence and Data Analytics (AIDA) Research Lab, College of Computer and Information Sciences, Prince Sultan University, Saudi Arabia
[g] Faculty of Computers and Information Technology, University of Tabuk, Tabuk, Saudi Arabia
[h] Faculty of Engineering, Mahallat Institute of Higher Education, Mahallat, Iran
[i] UCSI Graduate Business School, UCSI University, No. 1 Jalan Menara Gading, UCSI Heights, Cheras, Kuala Lumpur, 56000, Malaysia
[j] Centre for Global Sustainability Studies & School of Social Sciences, Universiti Sains Malaysia, Malaysia

A R T I C L E I N F O

Keywords:
Educational data mining
Multi-criteria collaborative filtering
Knowledge discovery
Text mining
Course choice decision
User' preferences

A B S T R A C T

Massive Open Online Courses (MOOCs) provide learners with high-quality and flexible online courses with no limitations regarding time and location. Detecting users' behavior in MOOCs is an important task for course recommendations. Collaborative Filtering (CF) is considered the widely approach in recommender systems to provide a online learner courses according to similar learners' preferences in an e-learning environment. The current research provides a novel framework through machine learning techniques to propose course recommendations in MOOCs according to the uses' preferences and behavior. The method is developed using multi-criteria ratings extracted from users' online reviews. We use Latent Dirichlet Allocation (LDA) for text mining, Decision Trees for decision rule generation, Self-Organizing Map (SOM) for users' reviews on courses and the fuzzy rule-based system for users' preferences prediction. We also adopt a feature selection method to select the most important criteria for users' preferences prediction. The method is evaluated using the data collected from an online learning platform, Udemy. The results showed that the method is able to accurately provide relevant courses to the users tailored to their preferences. The method has the potential to be implemented as a recommendation agent in the MOOC websites for course recommendations.

1. Introduction

Nowadays, multi-disciplinary methods along with the application of information as well as communication technologies (ICT) have caused significant changes in educational learning systems from conventional teaching toward modern learning without limitations of time and location (Gunga & Ricketts, 2008; Stošić & Stošić, 2015). Educational organizations are interested in employing ICT programs in their educational learning developments to promote the significance of learning and improve professional efficiency (Sakamoto, 2002; Stošić & Stošić, 2015). Prior research considering perceptions of students about online learning (Hara & Kling, 1999; Kim, Liu, & Bonk, 2005; Sloane, 1997; Smart & Cappel, 2006) have referred to different advantages, including convenient and flexible features, higher motivation for work, a better understanding of the material provided during the course, higher quality of learning, high accessibility, and providing instant feedback by the learners and the educators. According to these research findings, online teaching and learning processes make considerable impacts on the efficiency of higher education.

Data Mining or DM can be defined as the process through which beneficial information is extracted from big datasets using a combination of statistical as well as artificial intelligence methods (Wang, Sanín,

* Corresponding author.
E-mail address: nilashidotnet@hotmail.com (M. Nilashi).

https://doi.org/10.1016/j.eswa.2022.117092
Received 23 September 2020; Received in revised form 8 January 2022; Accepted 28 March 2022
Available online 4 April 2022

& Szczerbicki, 2013). The goal of data mining is to discover associations, patterns and stochastic structures from the available data. Educational Data Mining or EDM is among the evolving interdisciplinary research domains in education (Xing, Guo, Petakovic, & Goggins, 2015) and refers to the application of DM strategies to improve educational data in order to improve the education systems and address their shortcomings. EDM has currently evolved as a new research domain for scholars of various fields across the globe. Development of e-learning mechanisms, educational support, collecting educational data, and predicting the performance of the students are among applications of EDM. Moreover, EDM may be applied to develop e-learning mechanisms (Lara, Lizcano, Martínez, Pazos, & Riera, 2014), pedagogical support (Hung & Crooks, 2009), clustering educational data (Chakraborty, Chakma, & Mukherjee, 2016), and student performance predictions (Kabra & Bichkar, 2011).

Over the previous few years, innovations with the highest importance in the field of educational mechanisms have been associated with the establishment of novel technologies for education based on the web (Ha, Bae, & Park, 2000; Onan, 2020a; Wang, Xie, Au, Zou, & Wang, 2020). This procedure has become significantly important and a variety of web courses have been established in recent years. Nowadays, large-scale and open-access learning based on web is widely offered by Massive Open Online Courses (MOOCs). MOOCs are designed to be used by large numbers of learners with flexible features of learning in which advanced and high-quality educational materials can be delivered at large scale.

Educators can acquire feedback from the students' learning experiences in traditional teaching contexts through face-to-face interactions with students, making the continuous assessment of their teaching possible (Sheard, Ceddia, Hurst, & Tuovinen, 2003). Decision making regarding classroom processes includes observation of the students' behaviors, analysis of historical data, and estimation of the efficiency of educational methods. Nevertheless, when students work in electronic contexts, this informal control will be impossible; educators have to search new ways to obtain the required information. Organizations, through designing MOOCs to control distance education sites, can obtain significant amount of data which are produced by web servers in an automatic manner and collected in server access logs. Educational systems which have both adaptability and intelligent features are considered as solutions in providing richer sources to the learners. It is sought to provide the learners with individual teaching by considering their personal interests and priorities of preferences. Web-based learning context can record considerable learning behavioral characteristics of the students, and can therefore supply a great amount of learning profile. As web-based solution for online learning, MOOCs have showed effectiveness in pedagogical environments, however, most recent MOOCs for web-based courses have been according to static learning materials that do not consider the web personalization strategies. Accordingly, researchers are increasingly becoming interested in the automatic analysis of learners' interaction data through learning EDM tools.

Data mining or knowledge discovery in databases (KDD) can be defined as extracting implicit and attractive patterns automatically from big data sets. Its application is not only aimed at learning a model for the learning processes or students' modeling (Tang & McCalla, 2002), but also at examining and improving e-learning mechanisms through discovery of beneficial learning information from learning portfolios (Hwang, Chang, & Chen, 2004). Data mining can be mentioned as one step of the total KDD process which includes pre-process, data mining, and post-process. Thus, data mining is supposed to be useful for knowledge extraction from e-learning systems when the existing information is analyzed in the form of data produced by the users. Accordingly, discovering the patterns according to which the system is used by teachers as well as students will be the major goal along with discovery of the students' learning behavioral patterns as a critically important issue. Moreover, the educational data which has been created by the learners and educators will be used by KDD for the construction of the models which are subsequently employed for various aims, including predicting learners' preferences and their efficiency.

Many websites for online courses use sophisticated feedback system to help the learner in their learning decision making and expressing their experience on the quality of courses from different perspectives. As an example, Udemy provides many online programming courses along with a feedback system to get the users' comments based on their experiences from the courses. This website provides users with two types of electronic Word-of-Mouth (eWOM) information, namely aggregated rating and individual review. Users are encouraged by the MOOC to give ratings to the courses taken by them, while sharing their learning experiences with other users via learner reviews. Furthermore, an integrated rating is provided for every course according to the input given by the total reviewers which aims at presenting an overall assessment for the quality of the course. EWOM systems can be used more effectively compared to conventional WOM methods that are unable to classify overall opinions of the assessors and come to an overall assessment of the products. Since it is expected that overall ratings and individual reviews cooperate to make an effective online education, their effective application in MOOCs is possible with the aim of online learning system for the available courses while personalized recommending systems are designed.

The MOOC websites daily generate a huge amount of data from the educators and learners through their feedback system. The growing rate of the collected data stored in the databases is exponential. However, the stored data is kept by the system which is mainly used for statistical results. The stored data is valuable and should be effectively used for users' decision making, helping them to receive the more suitable resources tailored to their preference. As MOOCs with personalized recommendation systems can help the users save the time in finding the online courses (Campos, dos Santos, & Oliveira, 2020; Huang & Lu, 2018; Li et al., 2020; Ouertani & Alawadh, 2017), they will be more convenient in relation to the traditional MOOCs. In addition, a MOOC with a personalized recommender system would be useful, not only for saving time, but also because it can provide novel and diverse list of courses to the individuals (Xia, 2019; Zhang et al., 2019).

To the best of our knowledge, this research is the first one that focuses on the use of eWOM and individual reviews to investigate their effectiveness in MOOCs through designing a personalized recommender system. Specifically, this study for the first time is developed to use eWOM for developing a multi-criteria collaborative filtering which is believed to efficiently enhance the MOOCs performance for online course recommendations tailored to the learners' preferences. Overall, the contributions of this research are as follows:

i. The present research aimed at developing an online recommending system to assist MOOC websites to guide the users in their choice decision of the most suitable online courses which is a significant contribution in intelligent education systems. A multi-criteria CF recommendation algorithm is used to develop this system. The quality of recommendation mainly relies on the quality of the collected data and the ratings gathered from the MOOC websites. As users' satisfaction may not be revealed based on only the overall ratings on the courses, the proposed method consider multi-criteria ratings which are discovered from their reviews on the online course. Specific dimensions of the courses may better present the learners' preferences in MOOCs. In addition, degree of the users' experiences can be more conveniently evaluated by the multi-criteria rating scheme. The advantages of multi-criteria ratings on the recommendation results has been provided by (Adomavicius & Kwon, 2007).

ii. Several research works have determined the course ratings on a value scale; as an instance, a scale of 1 to 5 is specified in which 1 represents the lowest priority and 5 represents the highest for the desired course by a user. Some preference models have also been

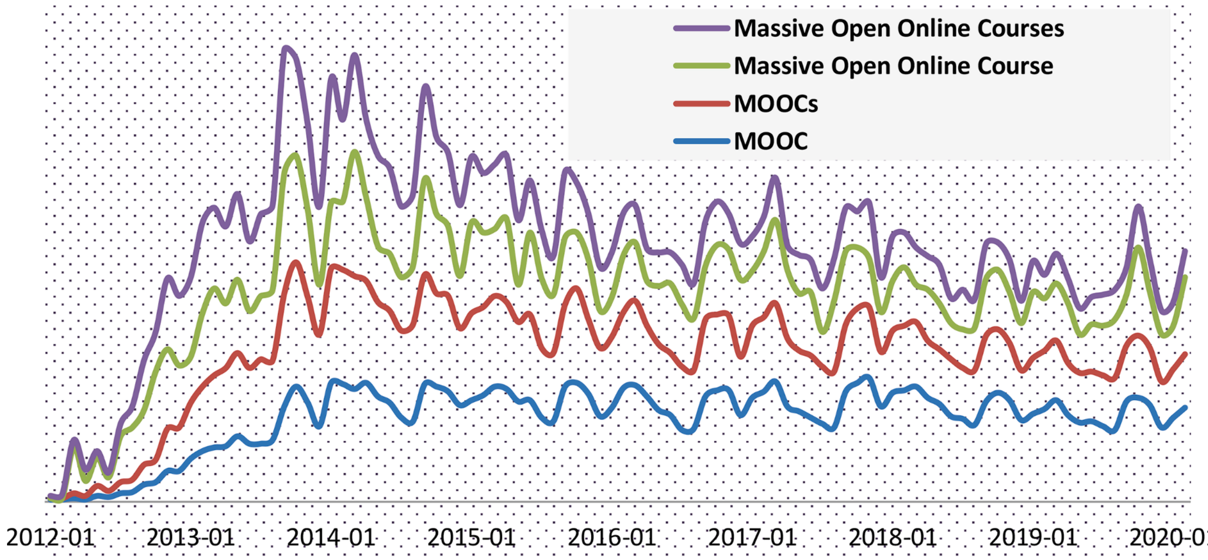


Fig. 1. Google Trends for MOOC related keywords.

proposed in special application domains which the learners prefer to show their priorities using linguistic terms, including "really interested" or "not interested" for the features of a course. Consequently, recommendations to the online users can be made accurately with the fuzzy logic approaches. This has attracted different scholars toward application of fuzzy set theory in recommending systems, so that higher accuracy and effectiveness of the recommendations could be achieved. Nevertheless, little research has been reported regarding application of a recommending system for MOOCs resource recommendations through fuzzy approaches.

iii. The method suggested in this study is based on multi-criteria CF. The crisp-based similarity algorithms are extended to fuzzy-based similarity algorithms in multi-criteria CF. The fuzzy set theory has become a lending procedure in the recommendation systems. Fuzzy set theory can address the problems in recommendation methods which are uncertain in nature (Cornelis, Lu, Guo, & Zhang, 2007). In particular, the proposed method is capable of correctly tackling linguistic variables employed for the explanation of users' personal demands, while recommending the educational resources according to uncertain data (Zhang et al., 2013).

iv. CART which is a prediction machine learning approach is applied to generate membership functions as well as fuzzy rules for the development of the suggested algorithms. Prediction of the overall course' ratings for recommendations will be performed using fuzzy rules. As a non-parametric technique, CART is used to develop decision rules, which can detect non-linear associations of the explanatory variables to assist in the discovery of main patterns from complicated datasets. This technique is robust to outliers. Typically, outliers are isolated in individual node or nodes by the splitting algorithm in CART.

The remainder of this paper is organized as follows. Literature review is presented in Section 2. We present the method in Section 3. Section 4 provides the method evaluation results. In Section 5, we present the discussions on the results. Finally, the work is concluded in Section 6.

2. Literature review

2.1. MOOCs and eWOM

The emergence of the internet has led to substantial changes in social behaviors which can be observed in the development of online learning along with Massive Open Online Courses (MOOCs) which have been effective tools in learning processes (Onan & Toçoğlu, 2020). There are many advantages for online learning in terms of reducing costs and flexibility. MOOCs are growing and reflect an exclusive promotion to the learning context, representing one of the most novel forms of open educational resources development. MOOCs indicate a really exclusive addition to the learning context. Since 2012, a growing attention has been observed for the use of MOOCs as witnessed in Google trends (see Fig. 1).

One of the critical factors in consumers' decision to purchase is electronic word of mouth or e-WOM, on which users are recently relying to communicate across the globe (Nilashi et al., 2019b). The increasing growth of the amount of information shared on the internet makes access to the required information difficult and leads a significant number of users toward optional reading habits. Research shows that a lot of users just consider the first five search pages, leading to under-representation of a significant number of business entities in the search results (Ahani et al., 2019). According to evidence, filtering e-WOM through computers can enable the users in finding the most relevant information in the least time and the lowest amount of efforts. Several systems including Amazon's book recommendations, Netflix's movies as well as YouTube's video suggestions have already indicated the realization of this concept. In addition, different applications such as blogs and social networks have been resulted from the fast expansion of web-based technology, helping to facilitate the employment of social network and subsequently obtain online resources. The fast development of User-Generated Content or UGC has faced the users with difficulties in efficient extraction of related content from the archives, particularly to make appropriate decisions.

Currently, much attention has been paid to online review in education websites including Udemy and Coursera that focus on online courses ratings. Comprehensive discussion channels, including online reviews or recommendations are changing into sources of real perspectives and queries with emergence of fast growing web 2.0 communication channels. The websites are frequently visited by the users in the search of appropriate online courses. Yet, an enormous quantity of information is provided, challenging them to select the most appropriate courses. Big data analytics can assist the users in such conditions, since these progressed technologies can deal with the large population of online learners and bring about considerable profits for the educational entities using MOOCs.

2.2. MOOCs and recommendation systems

Recommendation agents propose significant values to the users as well as business entities as a motivating market technology (Ahani et al., 2019). This way, the users are assisted in obtaining knowledge about a special product or service via big selection sets, while simultaneously businesses take advantage changing the browsers into shoppers, improving cross-sales and enhancing users' loyalty through provision of customized browse experiences. Different industrial entities have used these systems to deal with the challenge of information overload. The recommending systems include technologies which filter personalized information and are capable of filtering items according to the users' priorities, and have therefore the potential of predicting probable preferences (Nilashi et al., 2019b). The recommending systems are effective because they can obtain information on the users' favorite items through analysis of their prior behaviors, which means that the systems learn about the users' preferences from their prior purchase and can consequently identify their future preferred items.

Classification of these systems is performed according to their functions as collaborative filtering, context-based and hybrid techniques. The first type takes the interest profile of an active user into account with the profiles of those who have the same favorites. The major drawback of this system is associated with the requirement of more information of the consumers to recommend; therefore, they do not show enough effectiveness when there is a new user or a new item. The second type of recommendation agents adapts the recommendations to the recent conditions of the users and requires a suitable matching of the users' preferences with contextual components. Eventually, hybrid systems use a combination of collaborative and context-based methods.

Since the recommendations of the CF systems are generated through a two-step procedure, its overall success is closely associated with the phase of neighborhood formation that is dependent on the previous preferences of the users (Nilashi et al., 2019b). Each user's accounts include only one preference value for every item which does not consider the behavior of the users for the item' features. This issue in the modeling of recommendation system has been discussed by (Adomavicius & Kwon, 2007), while a novel approach has been proposed, according to which items are evaluated through their features. This approach aimed to improve the recommendation precision through multi-perspective ratings in relation to the single-ratings. CF systems which use these multi-criteria data are known as multi-criteria CF systems which present a promising area in the context of recommendation agents for future studies (Nilashi & bin Ibrahim, O., and Ithnin, N., 2014).

At the age of big data, collection, storage and analysis of the data provided by MOOCs is valuable for the educational systems. The quality of courses can be improved and appropriate targeted marketing attempts will be made using these data. Given the persistent and considerable flow of data, its accessibility and analysis in real-time conditions helps learners to make decisions, while course designers as well as online education systems are assisted to improve the provided courses. Recommender systems have been effective in handling large data in the educational systems (Khalid, Lundqvist, & Yates, 2020). Several attempts have been made on the development of different methods for educational resources recommendations (Assami, Daoudi, & Ajhoun, 2018; Bousbahi & Chorfi, 2015; Fu, Liu, Zhang, & Wang, 2015; Khalid, Lundqvist, Yates, & Ghzanfar, 2021; Prabhakar, Spanakis, & Zaiane, 2017; Sebbaq, el Faddouli, & Bennani, 2020; Wang, Xie, Wang, Lee, & Au, 2021). Although these studies have tried to solve several issues of educational resources recommendations by proposing methods by different machine learning approaches, however, the usefulness of multi-criteria ratings in MOOCs recommendation systems is not investigated. Accordingly, this study aims to use machine learning techniques for the development of a new recommender system for MOOCs based on multi-criteria ratings.

2.3. Related work on MOOCs and learning analytics

There have been several studies on MOOCs and improvements of these systems using different approaches. In this section, we discuss on the previous works and provide their results.

Mukala, Buijs, and Van Der Aalst (2015) employed multiple process mining approaches to examine a Coursera MOOC dataset and provided some indications in terms of helpful insights and guidance that could inspire intervention strategies to improve Coursera MOOC' quality and delivery. Van den Beemt, Buijs, and Van der Aalst (2018) focused on the analysis of structured learning behaviour in MOOCs. They proposed an approach based on process mining and clustering. They discovered four main groups of students, each indicating a distinct pattern of behavior ranging from only beginning to thoroughly finish the course. They extracted the event log of 16,224 students to discover the patterns. The results of their study revelated that successful students have a more consistent learning behavior, according to process mining methodologies. Abdelali (2016) focused on EDM and performed MOOCs videos mining using a metadata-based approach. They employed the clustering technique to provide learners with better search results. In the context of MOOCs, Tang, Xing, and Pei (2018) investigated the temporal dimension of forum participation using educational data mining techniques. The results revealed three main segments with dissimilar longitudinal participation trajectories. They found that the analysis of longitudinal forum participation is more effective in differentiating learner performance compared to the numerically aggregated measure. Liang, Yang, Wu, Li, and Zheng (2016) used logistics regression, support vector machine, random forest and gradient boosting decision tree for dropout prediction in Edx MOOCs. They collected 39 courses data from XuetangX platform. They achieved the highest prediction accuracy using a gradient boosting decision tree. In the context of MOOCs, Peral, Maté, and Marco (2017) focused on the application of data mining in identifying relevant key performance indicators. They proposed a new approach and used two case studies one on MOOC courses and another on Open Data from the education sector. Yousef, Chatti, Wosnitza, and Schroeder (2015) investigated the role of clustering in MOOCs data. They used a clustering approach to analyze different objectives of MOOC stakeholders to better understand the learners' behavior. The clustering approach discovered eight segments (i.e., flexibility, high-quality content, lifelong learning, blended learning, instructional design and learning methodologies, student-centered learning, and openness, network learning) from the data. Boroujeni and Dillenbourg (2019) proposed two methods to extract the patterns from activity sequences. The predefined patterns were extracted from learners' interactions with the course materials. They used a clustering pipeline for modelling and clustering activity sequences. de Barba et al. (2020) examined session behavioural data in MOOCs using clustering and group comparison tests. They found that students who demonstrated higher levels of effort regulation and time management received longer and more sessions throughout the course. Liao, Tang, and Zhao (2019) presented a method for course drop-out prediction in MOOCs using tensor completion and clustering approaches. A local tensor was used to cluster the MOOCs data and a high-accuracy low-rank tensor completion was used for drop-out prediction. Khalil and Ebner (2017) performed the clustering of patterns of engagement in MOOCs using learning analytics techniques. In their study the students were classified into appropriate categories based on their level of engagement. Albahr, Che, and Albahar (2021) developed a novel approach to extract keyphrase from MOOC video lectures. Their method was based on an unsupervised cluster-based

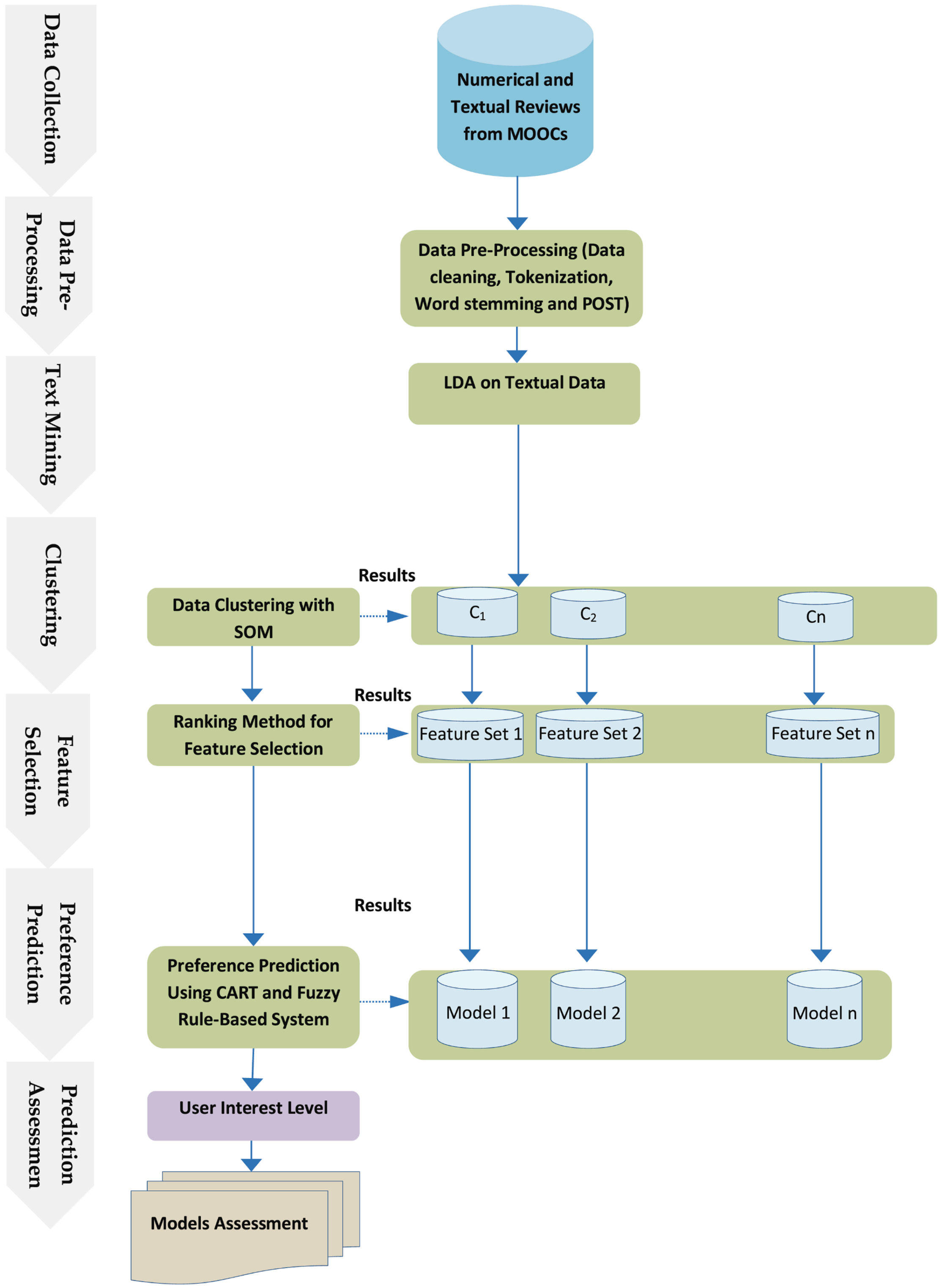


Fig. 2. Proposed method for multi-criteria CF in MOOCs.

approach, SemKeyphrase, for keyphrase extraction from MOOC video lectures.

Although significant contributions have been made in the previous studies on discovering patterns from MOOCs data, it is found that there is a lack of study in developing recommendation systems in MOOCs context through the use of machine learning techniques and multi-criteria collaborative approach. In addition, instead of using crisp approaches, the use of fuzzy set theory with the aid of CART can be a new contribution in the context of MOOCs. Fuzzy logic approaches have shown to be effective in modeling human behavior. In the context of MOOCs, fuzzy logic with the aid of machine learning techniques may better reveal the learner' preference from a set of real-world data. By developing a multi-criteria recommendation system, this study is therefore conducted to investigate the effectiveness of combining these techniques in the context of MOOCs. Lin et al. (2021) focused on an adaptive course recommendation in MOOCs. They designed a dynamic attention mechanism to track the changes in users' preferences.

3. Methodology

3.1. Preliminaries on fuzzy techniques

Several fundamental concepts of fuzzy sets are first introduced in order to describe the suggested method and obtain the associated theorems. The concepts are employed to calculate the linguistic term similarities and predict the ratings in the suggested recommendation process (Zadeh, 1965).

Definition 1. For a fuzzy number $\widetilde{a}$, the λ-cut is defined as:

$$\widetilde{a}_{\lambda} = \{x, \mu_{\widetilde{a}}(x) \geqslant \lambda, x \in R\} \tag{1}$$

where $\mu_{\widetilde{A}}(x)$, $x \in [0,1]$ in X, is the Membership Function (MF) in fuzzy set $\widetilde{A}$, $\widetilde{a}_{\lambda} = [a_{\lambda}^{-}, a_{\lambda}^{+}]$ is a nonempty bounded closed interval contained in a universe of discourse X which a_{λ}^{-} is the lower bounds of the closed interval and a_{λ}^{+} is the closed interval' upper bounds.

Definition 2. Suppose $\widetilde{a}$ is a TFN (Triangular Fuzzy Number). It is defined by a triplet (a_0^{-}, a, a_0^{+}). Accordingly, the MF $\mu_{\widetilde{a}}(x)$ is defined as:

$$\mu_{\widetilde{a}}(x) = \begin{cases} 0, & X < a_0^{-} \\ \dfrac{x - a_0^{-}}{a - a_0^{-}}, & a_0^{-} \le X \le a \\ \dfrac{a_0^{+} - x}{a_0^{+} - a}, & a \le X \le a_0^{+} \\ 0, & a_0^{+} < X \end{cases} \tag{2}$$

Definition 3. For any $\widetilde{a}, \widetilde{b} \in F_{+}^{*}(R)$, the set of all finite PFN (Positive Fuzzy Numbers) on R, and $0 < \delta \in R$, we have:

$$\widetilde{a} + \widetilde{b} = \bigcup_{\lambda \in [0,1]} \lambda [a_{\lambda}^{-} + b_{\lambda}^{-}, a_{\lambda}^{+} + b_{\lambda}^{+}] \tag{3}$$

$$\delta \widetilde{a} = \bigcup_{\lambda \in [0,1]} \lambda [\delta a_{\lambda}^{-}, \delta a_{\lambda}^{+}] \tag{4}$$

$$\widetilde{a} \times \widetilde{b} = \bigcup_{\lambda \in [0,1]} \lambda [a_{\lambda}^{-} \times b_{\lambda}^{-}, a_{\lambda}^{+} \times b_{\lambda}^{+}] \tag{5}$$

Definition 4. Suppose $\widetilde{a}$ and $\widetilde{b}$ are two fuzzy numbers. Then $\widetilde{a} = \widetilde{b}$ if $a_{\lambda}^{-} = b_{\lambda}^{-}$ and $a_{\lambda}^{+} = b_{\lambda}^{+}$ for any $\lambda \in [0,1]$.

3.2. Proposed method for recommender system in MOOCs

Fig. 2 shows the suggested model that has employed supervised learning, text mining and clustering machine approaches. According to the figure, text mining, clustering, CART as well as fuzzy rules can be considered as main components of the proposed method. We use LDA to analysis textual reviews on the courses. CART is taken into account in the construction of the decision rules for the techniques based on fuzzy rules to learn the prediction models. The decision rules can be used by the system to make course prediction according to the users' preferences. CART has been known as a supervised machine learning strategy whose rules are simply understood in linguistic forms. CART is used because database of rules is an important component of the fuzzy rule-based methods which require the decision rules to perform predictions. This method is capable of generating the rules automatically and does not require intervention by humans. Therefore, CART has been employed in the present work for automatic generation of decision rules to be employed in fuzzy rule-based method.

Data collection is initially done from a MOOC website, Udemy.com. Next, pre-processing of the collected data is carried out. The data gathered form the educational context needs initial pre-processing in order to change it to a suitable format for data mining. There are different pre-processing functions, the most important of which include cleaning, attributes selection, attribute transformation, and data integration. Then, discovery of criteria from online text reviews was performed using LDA. In the following stage, clustering is done by the use of SOM, and then CART is applied in every cluster to discover the decision rules. Eventually, construction of the prediction models is completed using fuzzy rule-based method.

Learning of the prediction functions for the items as well as users has been taken into account in the present work similar to the prior works on multi-criteria CF. A weighted approach has been used in every cluster to obtain a combination of the prediction functions. According to previous studies (Jannach, Karakaya, & Gedikli, 2012; Nilashi, Jannach, & bin Ibrahim, O., and Ithnin, N., 2015), we use Eq. (6) as a general weighting scheme, which is defined as:

$$r_{u,i} = w_u{}^{*} r_{u,i}^{user} + w_i{}^{*} r_{u,i}^{item} \tag{6}$$

In Eq. (6), w_u and w_i are the weight of r^{user} andr^{item}, respectively. According to the definitions of fuzzy set, we extend the above prediction method for fuzzy-based prediction.

$$\widehat{r}_{u,i} = w_u \times \widehat{r}_{x,u}^{user} + w_x \times \widehat{r}_{x,u}^{item} = \bigcup_{\alpha \in [0,1]} \alpha \left[w_u \times a^{user} \widehat{r}_{x,u}^{-} + w_i \times a^{item} \widehat{r}_{x,u}^{-}, w_u \times a^{user} \widehat{r}_{x,u}^{+} + w_x \times a^{item} \widehat{r}_{x,u}^{+} \right]. \tag{7}$$

A major disadvantage of different CF recommending systems is sparsity that significantly can impact on the accuracy of items recommendation. CF algorithm generates inefficient recommendations when there are a lower number of ratings. Enough amounts of rating data are required by collaborative recommendation algorithms. Clustering as well as fuzzy rule-based methods have been used in the present work to deal with this problem. Moreover, the technique proposed by (Adomavicius & Kwon, 2007) was used in the present work to establish similarities of the users (see Eq. (8)), after which it was applied as a fuzzy-based similarity calculation technique (see Eq. (9)). The average similarity of the two users is achieved by the use of the suggested method according to Eq. (8) (Stošić & Stošić, 2015).

$$sim_{avg}(u, u') = \frac{1}{k+1} \sum_{c=0}^{k} sim_c(u, u'), \tag{8}$$

According to the definition based on fuzzy set, we define $sim_c(u, u')$ as:

$$sim_c(u,u') = \frac{\sum_{x\in I_{u,u'}} \int_0^1 \frac{1}{2}\left[\left(r^-_{x,u_\alpha} - \overline{r}^-_{u_\alpha}\right)\left(r^-_{x,u'_\alpha} - \overline{r}^-_{u_\alpha}\right) + \left(r^+_{x,u_\alpha} - \overline{r}^+_{u_\alpha}\right)\left(r^+_{x,u'_\alpha} - \overline{r}^+_{u_\alpha}\right)\right]d\alpha}{\sqrt{\sum_{x\in I_{u,u'}}\left(\int_0^1 \frac{1}{2}\left[\left(r^-_{x,u_\alpha} - \overline{r}^-_{u_\alpha}\right) + \left(r^+_{x,u_\alpha} - \overline{r}^+_{u_\alpha}\right)\right]d\alpha\right)^2} \times \sqrt{\sum_{x\in I_{u,u'}}\left(\int_0^1 \frac{1}{2}\left[\left(r^-_{x,u'_\alpha} - \overline{r}^-_{u_\alpha}\right) + \left(r^+_{x,u'_\alpha} - \overline{r}^+_{u_\alpha}\right)\right]d\alpha\right)^2}}. \tag{9}$$

The recommending system initially discovers the active or target users along with the active or target courses in this phase. The tasks of rating predictions as well as recommendations are carried out in the next stage. The first task includes every algorithm for the prediction of the ratings associated with a list of courses according to determined priorities of a specific active user. In the second phase, the system includes the ranking of a list of items which have not been rated for active users after which the top-K recommendations including the first K courses in the list of recommendations are provided.

3.2.1. CART algorithm

The CART classification procedure employs historical data for construction of decision trees which can be applied for classification or regression tasks, known as innovative algorithm in data mining (Nilashi, Ibrahim, Ahmadi, & Shahmoradi, 2017). Actually, CART can be defined as a binary division of the attributes space, which means that CART generates binary decision trees capable of separating scalar features from constant ones. There are two steps in each CART algorithm:

i. Generating the decision trees: generation of decision trees is performed according to the training dataset. The decision trees which have been produced need to be big enough;
ii. Pruning of the decision trees: pruning of the produced decision trees is carried out through verification dataset, after which the optimum sub-tree will be chosen. Meantime, the standard of pruning will be according to the minimum loss function.

Creating a decision tree in CART includes recursive construction of a binary decision tree. These decision trees are applicable to classify and perform regressions. However, the first application has been discussed in the present paper. In this application, the Gini coefficient minimization criterion has been used to select the attributes and create a binary tree. Algorithm 1 indicates the process of generating the decision trees in CART.

Algorithm 1. CART Algorithm.

INPUT: Training dataset of D along with conditions to stop computations
OUTPUT: Decision trees

i. Consider the training dataset of nodes as D and compute the Gini coefficient of the available attributes on the dataset. Accordingly, every attribute will have its own value a. Based on the sample test $A = a$ for "yes" or "no", partition D into D_1 and D_2, and compute the Gini coefficient in the case that $A = a$.
ii. Over all probable attributes A and all potential partition points a, the attribute having the least Gini coefficient along with its relative partition points can be chosen as the optimum attributes and the optimum segmentation points. Based on the optimum attribute as well as the optimum segmentation points, the current node generates two children nodes each of which receive the training dataset based on the attributes.
iii. Perform Stages (1) to (2) in a recursive manner for the mentioned children nodes up to the time that stop conditions are satisfied
iv. Produce the CART decision tree.

When the number of samples in a node is fewer than the existing threshold, the Gini coefficient of the sample set falls below the existing threshold which means that the samples are primarily from one class, or when no other attributes can be found, the stopping conditions of the algorithm are met.

3.2.2. SMOTe

Application of sampling strategies in use of imbalanced datasets includes changes in the quantities of class data to establish balanced class distributions (Guo, Liu, Chen, Sun, & Wang, 2019). Synthetic Minority Oversampling Technique or SMOTE is a robust over-sampling method which was provided by (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). Despite other techniques which use random oversampling of instances through duplicating, SMOTE method makes novel artificial samples by the use of knowledge of the neighbors surrounding every instance in the minority class. The proposed method is extensively applied with significant success in addressing the functions of imbalanced datasets. Accordingly, a totally distinguished approach is used for over-sampling of the minority class, operating on the attribute space instead of the data. Fundamentally, SMOTE has been introduced to address persistent or discrete variables. In this regard, k-nearest neighbor algorithm is used to identify the k nearest neighbors in a specific minority data instance in a neighborhood. In an imbalanced dataset, every minority class instance is taken and the amount of synthetic samples is created together with the k minority class nearest neighbors in order to over-sample the minority class. The parameter N indicates the amount of synthetic samples produced for each original minority instance, and the parameter k is associated with the nearest neighbors should be already defined. Normally, N = 200% amount of oversampled instances and five nearest neighbors can be used by individuals. As an instance, $N/100$ neighbors from five nearest neighbors may be selected, while one synthetic sample will be produced in every direction.

Particularly, the process for generation of synthetic instances includes several stages as follows: S_{min} indicates the set of minority class from $\subseteq$ S, for every sample of $x_i \in S_{min}$, look for its k nearest neighbors with the use of Euclidean distance. For generating a synthetic sample, choose one of the k-nearest neighbors in a random manner, and then compute an attribute difference between x_i as well as its neighbor. Moreover, multiply the attribute vector difference by a random number $\alpha \in [0,1]$, and eventually perform addition of the vector to x_i so that the synthetic sample x_{new} can be obtained.

$$x_{new} = x_i + (\widehat{x}_i - x_i) \times \alpha \tag{10}$$

$x_i \in S_{min}$ indicates an example of minority class in the primary population, $\widehat{x}_i$ represents one of the k-nearest neighbors of x_i, and $\alpha \in [0,1]$ indicates a real random number. Thus, the new synthetic sample according to Eq. (10) will be a data point across the line segment between x_i and the k nearest neighbor $\widehat{x}_i$ which has been chosen in a random manner. The synthetic samples lead classifiers toward creation of bigger and less specified decision areas for minority class. Algorithm 2 will be the pseudo code for SMOTE.

	x_1	…	x_l	…	x_L	$\sum$
y_1						
$\vdots$			$\vdots$			
y_k		…		…		n_k
$\vdots$			$\vdots$			
y_K						
$\sum$			n_l			n

Fig. 3. Feature selection using correlation through the association between target attributes $\{y_1,\cdots,y_k,\cdots y_K\}$ and predictors$\{x_1,\cdots,x_k,\cdots x_K\}$.

Algorithm 2. SMOTE Algorithm

Input: Minority data $D^{(t)} = \{x_i \in X\}$; $i = 1,2,\cdots X$; Number of minority cases (X); Number of nearest neighbors (k); SMOTE percentage (N)

For $i = 1,2,\cdots,T$ **Do**

i. Find the k- nearest (minority class) neighbors of x_i.

ii. N = N/100

While $\widehat{N} \neq 0$ **Do**

a. Selecting the K nearest neighbors, call this $\widehat{x_i}$.

b. Selecting a random number $\alpha \in [0,1]$.

c. $x_{new} = x_i + (\widehat{x_i} - x_i) \times \alpha$

d. $S \leftarrow x_{new}$

e. $\widehat{N} = \widehat{N} - 1$

End While

End For

Output: Return S

3.2.3. LDa

Textual data analysis has been applied in many researches (Onan, 2018, 2020b, 2020c, 2020d; Onan, Korukoğlu, & Bulut, 2016). The statistical model of Latent Dirichlet Allocation (LDA) is mainly applied to find the fundamental abstract topics in a set of documents or textual data (Blei, Ng, & Jordan, 2003). According to the presumption of "beg of words", LDA indicates a document in the form of a combination of latent topics where the topic represents a multinomial distribution across the works. Each document has a specific combination ration regarding the topics and every topic has a determined distribution of words for itself. Latent topics can be discovered by LDA using the big review data which is not structured. This technique helps in identification of the optimal number of topics, while labeling the topics and analysis of the differences along with the topics corresponding significance will be possible for various products.

According to the assumptions made by LDA, it is possible to generate topic by every document (the fundamental topic distribution), as description of the topics is carried out across the words. Regarding every document, the words are firstly created according to the random distribution over the topics. In the next step, one topic should be selected in a random manner for every word in the document (after finding the distribution in the prior stage) and one word should be selected from the word-distribution across the vocabulary in a random manner. It is supposed that every document indicates the topics in various rations according to (1), whereas the words come from the topics according to (2). The efficiency of the technique is dependent on the assumption that the words which co-occur will be potentially from semantic-associated topics. Moreover, words which have rare co-occurrence will be subject to formation of notions which are associated with other distinguished topics.

3.2.4. Feature selection through correlation based filter method

Features selection is a critical strategy used in solving dimensionality problems in machine learning through selection of related as well as non-redundant attributes (Nilashi, Ibrahim, & Ithnin, 2014). Higher scalability, reliability and accuracy can be obtained for machine learning algorithms through selection of effective features with facilitation of features selection. It can be also beneficial in predicting over data analysis processes through selection of closed as well as relevant attributes. Majority of these algorithms employ statistic measurements, including mutual information, correlation, as well as information gain measurement. According to evaluation scales, three overall processes can be mentioned for selection of features, including Filters, Wrappers and Embedded strategies (Lal, Chapelle, Weston, & Elisseeff, 2006). The first method explores the internal features of the data in an open-loop process in an independent way from the classifier design. On the other hand, the second and third methods select features through interaction with classifiers. Although the second method uses a classifier assumption in the close-loop search so that the optimum feature subset is found, the third method incorporates the search in the construct of the classifier. These two methods usually lead to more accurate results compared to the first one; however, the first method does not usually result in over-fitting and has been reported to be more economic from the computational perspective, especially in comparison with the embedded method. The present work has used filter method to select data from the datasets.

As presented in Fig. 3, the association between two variables is measured through the contingency table constructed by these variables which $Y\{y_1,\cdots,y_k,\cdots y_K\}$ is the target attribute and $X \in \{x_1,\cdots,x_k,\cdots x_K\}$ is the predictor.

Considering $p_{kl} = \frac{n_{kl}}{n}, p_k = \frac{n_k}{n}$ and $p_l = \frac{n_l}{n}$, the mutual information between two variables (mutual dependence) is obtained as follows:

$$D(X,Y) = \sum_k \sum_l p_{kl} \times log_2 \frac{p_{kl}}{p_k \times p_l} \tag{11}$$

$$\mathrm{M}(Y) = -\sum_l p_k \times log_2 p_k \tag{12}$$

Accordingly, the symmetrical uncertainty is presented as follows:

$$p_{y,x} = 2 \times \left[\frac{D(X,Y)}{\mathrm{M}(Y) + \mathrm{M}(X)}\right] \tag{13}$$

The normal approximation of the ρ is used to evaluate the significance of the association as follows:

$$\mathrm{M}(X,Y) = -\sum_k \sum_l p_{kl} \times log_2 p_{kl} \tag{14}$$

The standard error can be obtained from:

$$\sigma_p^2 = 4 \times \sum_k \sum_l \frac{n_{kl}\left[\mathrm{M}(X,Y) \times log_2\left(\frac{n_k \times n_l}{n^2}\right) - [\mathrm{M}(Y) + \mathrm{M}(X)] \times log_2\left(\frac{n_{kl}}{n}\right)\right]^2}{n^2 \times [\mathrm{M}(Y) + \mathrm{M}(X)]^4} \tag{15}$$

The standard error under the null hypothesis is then computed as:

$$\sigma_p^2 = 4 \times \frac{\sum_k \sum_l n_{kl}\left[log_2\left(\frac{n_k \times n_l}{n \times n_{lk}}\right)\right]^2 - \frac{[\mathrm{M}(Y) + \mathrm{M}(X) - \mathrm{M}(Y,X)]}{n}}{n^2 \times [\mathrm{M}(Y) + \mathrm{M}(X)]^2} \tag{16}$$

where the critical region with the quantile of the Gaussian

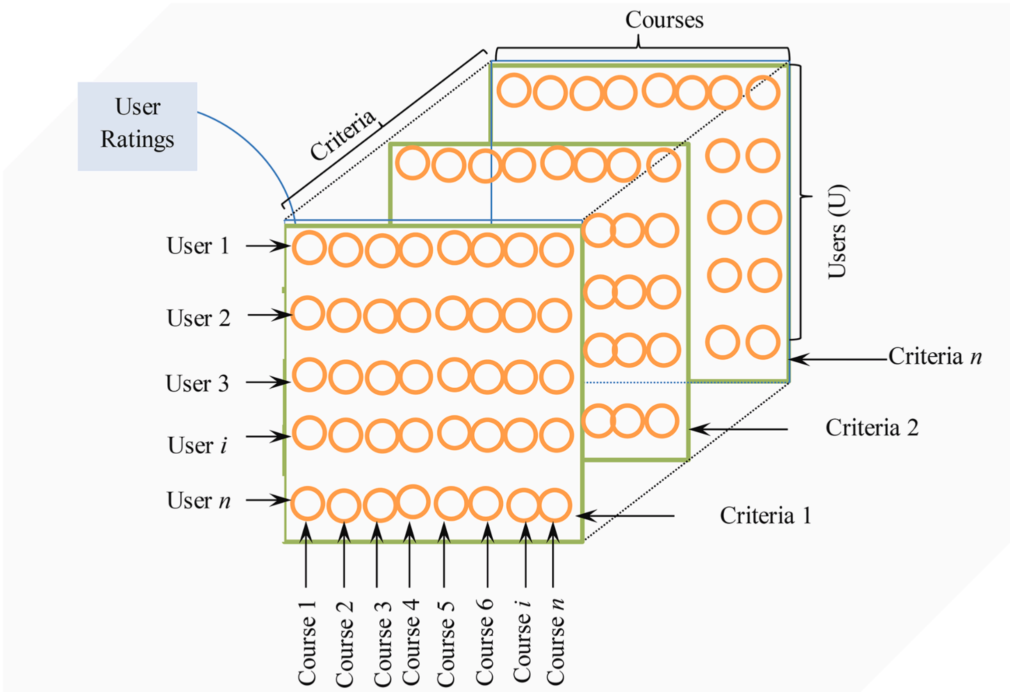


Fig. 4. Tensor of users' rating on courses' features.

Table 1
Descriptive statistics of collected data.

Assessment Criteria	Minimum	Maximum	Mean	Std. Deviation
User Overall Interest	0.000006	0.999690	0.496736	0.288080
Knowledge of Instructor	0.000116	0.999856	0.498117	0.290398
Value of Information	0.000059	0.999855	0.499211	0.290345
Clarity of Explanation	0.000071	0.999948	0.502938	0.288449
Engaging Delivery of Information	0.000006	0.999916	0.496196	0.287639
Accuracy of Course Description	0.000063	0.999958	0.501913	0.289031
Helpfulness of Practice Activities	0.000006	0.999690	0.496736	0.288080

distribution $u_{1-\alpha}$ is defined as follows:

$$\frac{\rho}{\sigma_p(0)} > u_{1-\alpha} \tag{17}$$

In the ranking approach, the correlation of each input variable with the output is calculated and the redundancies between the predictors are not considered. Finally, we rank the selected features in a decreasing order, according to the calculated ρ values.

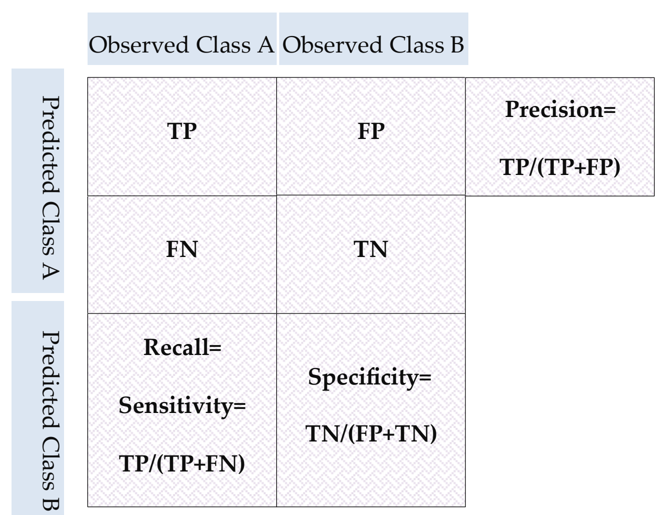


Fig. 5. Confusion matrix.

3.2.5. Self-Organizing maps method

Clustering approach is found to be effective in data analysis to improve the efficiency of the machine learning methods (Onan, 2019; Onan, Korukoğlu, & Bulut, 2017). This technique which is also called SOM is comparatively novel for educational recommending components. This non-linear technique includes an artificial neural network according to unsupervised learning, in which high-dimensional input data is mapped onto a regular, low-dimensional array which is typically two-dimensional (Ahani et al., 2019). Arrangement of the neurons in SOM is in a two-dimensional array for the generation of two-dimensional Kohonen maps to preserve data similarities (Nilashi *et al.*, 2019a). This algorithm focuses on representing the samples on a low-dimensional network or map, while the samples are kept in the farthest probable corresponding distance. Accordingly, training of the map is achieved by iterative updates of a set of weights allocated for every grid unit. Selection of sample vectors is performed across the training procedure, while the grid unit having the highest similarity of the weight vector will be identified. Update of the winning grid and neighbors across a special neighborhood domain will be then carried out using Eq. (18), in which w_k represents the weight vector for unit k, x_t indicates the recent sample, α_t indicates the learning rate and n_t represents the neighbourhood weight for iteration t.

$$w_t = w_k + \alpha_t n_t (x_t - w_k) \tag{18}$$

There is gradual decrease of the learning rate, neighborhood scope and weight with t to ensure locating of learning for every sample in a domain of the grid according to similarities with other samples. When the training is finished, mapping of the samples takes place based on

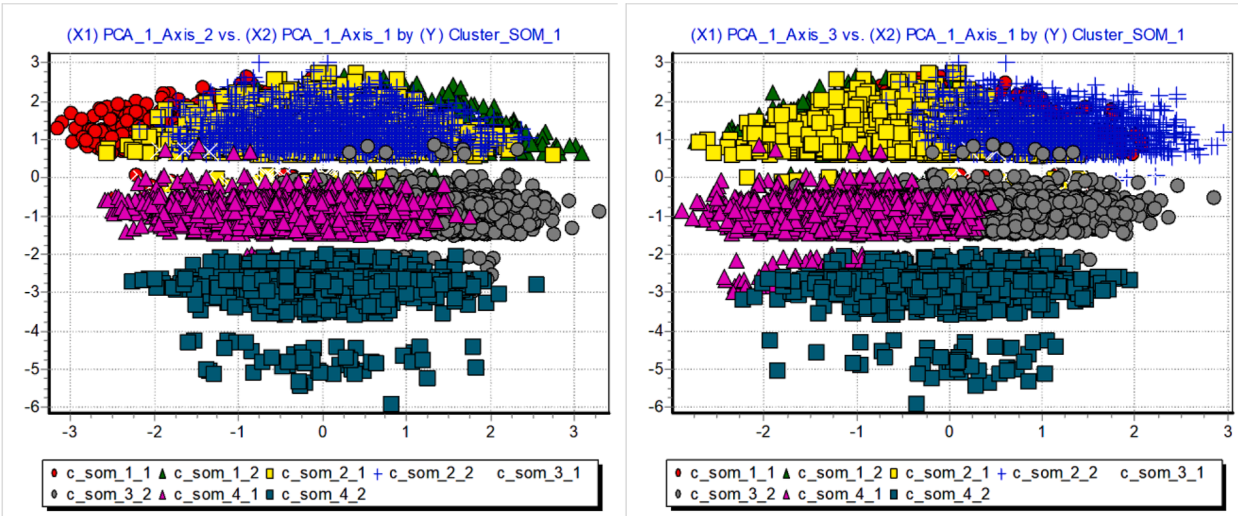


Fig. 6. Visualization of SOM clustering.

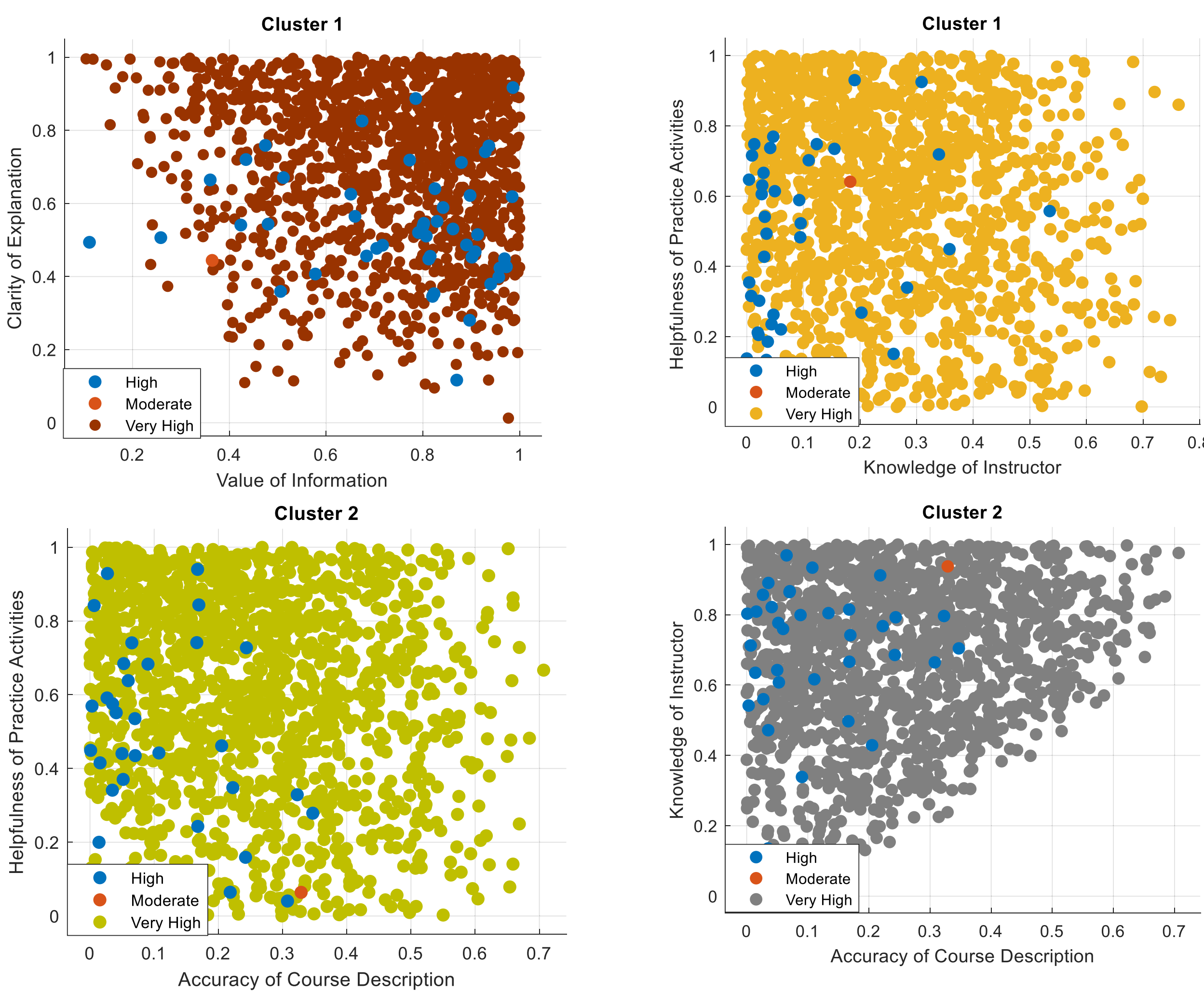


Fig. 7a. Visualization of SOM clustering based on discovered criteria.

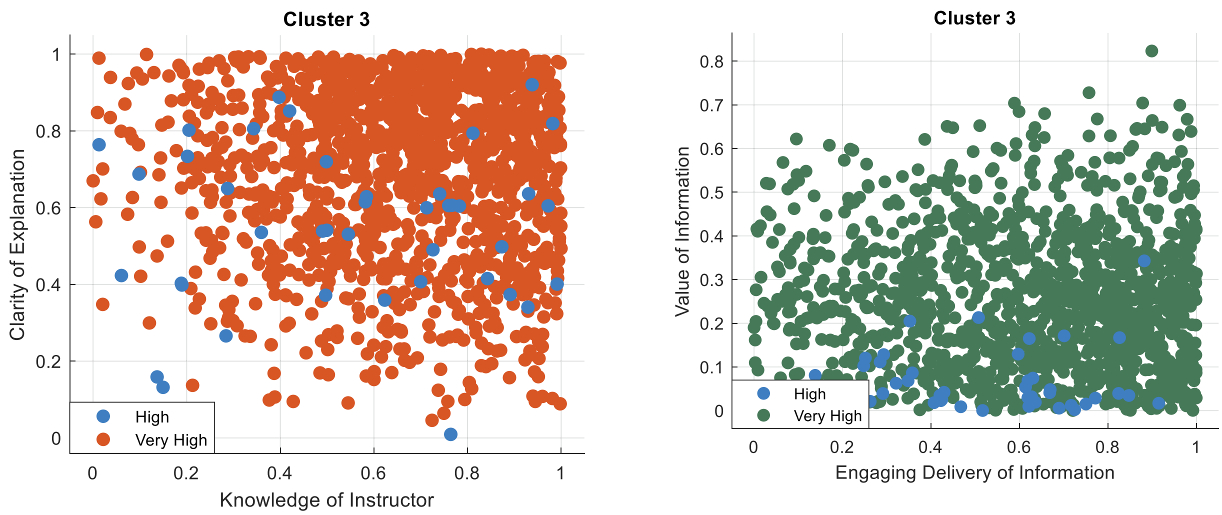


Fig. 7b. Visualization of SOM clustering based on discovered criteria (Cont.).

Table 2
SOM clustering results.

Criteria	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
User Overall Interest	4.975	4.985	4.982	4.999	4.035	4.029	4.016	2.988
Knowledge of Instructor	0.260	0.687	0.654	0.720	0.482	0.427	0.348	0.289
Value of Information	0.718	0.696	0.247	0.711	0.488	0.385	0.371	0.280
Clarity of Explanation	0.707	0.680	0.686	0.337	0.250	0.314	0.751	0.268
Engaging Delivery of Information	0.548	0.626	0.626	0.572	0.246	0.722	0.288	0.281
Accuracy of Course Description	0.706	0.250	0.676	0.729	0.589	0.332	0.353	0.272
Helpfulness of Practice Activities	0.585	0.602	0.645	0.555	0.477	0.380	0.422	0.285

Table 3
Feature selection in 8 clusters.

Criteria	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Knowledge of Instructor	√		√	√	√	√	√	√
Value of Information	√	√	√	√	√		√	
Clarity of Explanation	√	√	√			√	√	√
Engaging Delivery of Information	√	√	√	√	√	√	√	√
Accuracy of Course Description	√	√	√		√		√	
Helpfulness of Practice Activities	√	√	√	√	√	√	√	√

similarities with the grid unit weight vectors, after which their graphical representation will be performed. Moreover, the method is applicable to plot the new data on the maps of training datasets in order to be used in predictions.

4. Method evaluation

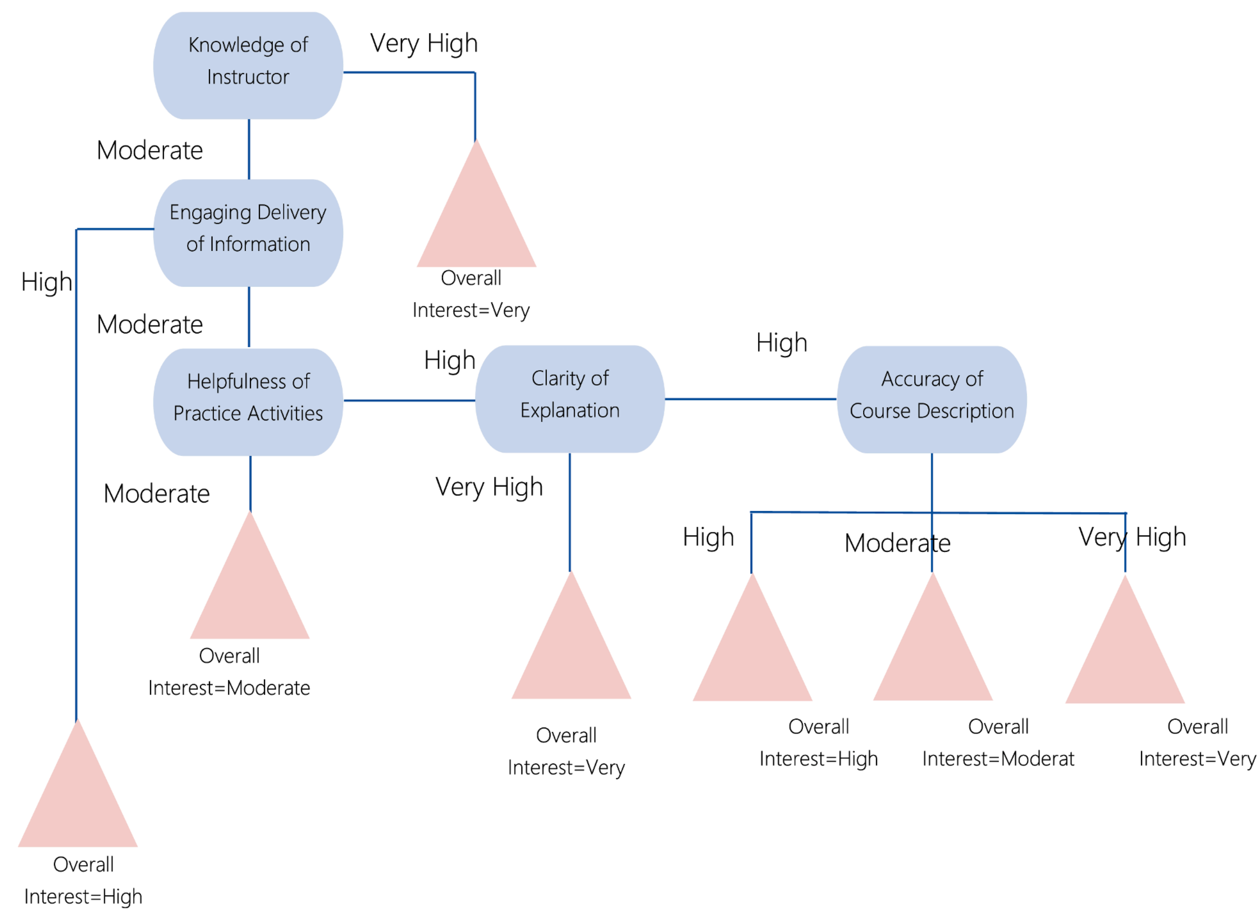
4.1. Data collection

Data collection was collected from Udemy at www.udemy.com during January 2020 and March 2020. The users can provide their reviews on Udemy' products based on User Overall Interest, Knowledge of Instructor, Value of Information, Clarity of Explanation, Engaging Delivery of Information, Accuracy of Course Description, and Helpfulness of Practice Activities. A total number of 11,256 ratings were obtained, including 4882 users and 2872 items. After topic mining by LDA for Knowledge of Instructor, Value of Information, Clarity of Explanation, Engaging Delivery of Information, Accuracy of Course Description, and Helpfulness of Practice Activities, the data was stored in the tensors as shown in Fig. 4. Descriptive statistics of collected data is presented in Table 1.

Every data source was preprocessed individually prior to the application of EDM methods to address two considerable problems which can repeatedly occur in educational data. The first problem is associated with high dimensionality which means a great number of features. This problem can challenge the prediction algorithms in achieving the desired results in a shorter time. The second problem is related to unbalanced data. In the case of significantly bigger number of instances in one class compared to the others, prediction algorithms may concentrate on the former instead of the latter. As the data sources under analysis have a lot of features to be addressed, they have the dimensionality problem. Therefore, experimental evaluation of the feature selection algorithms was performed for every data source. Furthermore, based on the data kept in the data source, there is no balance for the number of positive classes. Therefore, experimental evaluation of the datasets was carried out by the use of SMOTE algorithm.

4.2. Clustering, CART and fuzzy logic analysis

This research employs SOM to cluster the numerical ratings provided by the users on the different criteria (Knowledge of Instructor, Value of Information, Clarity of Explanation, Engaging Delivery of Information, Accuracy of Course Description, and Helpfulness of Practice Activities) of educational resources. For SOM clustering, we considered SOM2 × 2



- IF Knowledge of Instructor =Moderate
  - AND Engaging Delivery of Information =Moderate
    - AND Helpfulness of Practice Activities=Moderate **THEN** User Overall Interest = Moderate
    - AND Helpfulness of Practice Activities =High
      - AND Clarity of Explanation =High
        - AND Accuracy of Course Description =High **THEN** User Overall Interest = High
        - AND Accuracy of Course Description = Very High **THEN** User Overall Interest = Very High
        - AND Accuracy of Course Description = Moderate **THEN** User Overall Interest = Moderate
      - AND Clarity of Explanation =Very High **THEN** User Overall Interest = Very High
  - AND Engaging Delivery of Information = High **THEN** User Overall Interest = High
- IF Knowledge of Instructor = Very High **THEN** User Overall Interest = Very High

Fig. 8. A part of decision rules by CART in Cluster 1.

for 4 clusters, SOM2 × 4 for 8 clusters, SOM3 × 2 for 8 clusters, SOM3 × 3 for 9 clusters and SOM 4 × 4 for 12 clusters. The best clustering quality was obtained for SOM4 × 4 (see Fig. 6 and Fig 7). SOM clustering results for the clusters centroids are presented in Table 2. The feature selection method was used to find the most important features in each cluster. The results for feature selection are presented in Table 3.

The CART technique was then applied on the clusters to find the relationships among the criteria (discovered by LDA) for overall interest prediction. In CART, the relationships among the criteria were defined by the discovered decision rules. For output (User Overall Interest), we considered three labels as Low, Moderate and High. As an example, the first decision rule is defined as: Rule 1: IF “Knowledge of Instructor” is [Moderate] AND “Engaging Delivery of Information” is [Moderate] AND “Helpfulness of Practice Activities” is [Moderate] THEN “Overall Interest” = Moderate. ‘In addition, the second decision rule is: Rule 2: IF “Knowledge of Instructor” is [Moderate] AND “Engaging Delivery of Information” is [High] THEN “Overall Interest” = High. We present a part of decision rules in Figs. 8-10. We use these decision rues in fuzzy rule-based system to predict users’ overall preferences in Udemy.

MFs were employed for the users’ preference modeling in the present study, so that the prediction model could be constructed in fuzzy logic. Different kinds of MFs are available to provide systems according to theories of fuzzy sets. The most important MFs are Triangular, Generalized Bell-Shaped, Gaussian as well as P-Shaped. Triangular MF was

- IF Engaging Delivery of Information=High
  - AND Clarity of Explanation =High
    - AND Helpfulness of Practice Activities =High
      - AND Accuracy of Course Description =Moderate THEN User Overall Interest = Moderate
      - AND Accuracy of Course Description =High THEN User Overall Interest = High
      - AND Accuracy of Course Description =Very High THEN User Overall Interest = Very High
    - AND Helpfulness of Practice Activities =Moderate THEN User Overall Interest = High
  - AND Clarity of Explanation =Moderate THEN User Overall Interest = Moderate
- IF Engaging Delivery of Information =Moderate THEN User Overall Interest = Moderate
- IF Engaging Delivery of Information =Very High THEN User Overall Interest = Very High

Fig. 9. A part of decision rules by CART in Cluster 2.

- IF Value of Information =Low
  - AND Helpfulness of Practice Activities =Moderate THEN User Overall Interest = Low
  - AND Helpfulness of Practice Activities =High
      - AND Knowledge of Instructor=High THEN User Overall Interest = Moderate
      - AND Knowledge of Instructor =Low
        - AND Clarity of Explanation =Low THEN User Overall Interest = Low
        - AND Clarity of Explanation =Moderate
          - AND Engaging Delivery of Information =Moderate THEN User Overall Interest = Moderate
          - AND Engaging Delivery of Information =Low THEN User Overall Interest = Low
- IF Value of Information =Moderate
  - AND Knowledge of Instructor =High THEN User Overall Interest = Moderate
  - AND Knowledge of Instructor=Moderate THEN User Overall Interest = Moderate
- IF Value of Information =Very High
  - AND Knowledge of Instructor =High THEN User Overall Interest = Very High
  - AND Knowledge of Instructor=Moderate THEN User Overall Interest = High
- IF Value of Information =Moderate THEN User Overall Interest = Moderate

Fig. 10. A part of decision rules by CART in Cluster 3.

used in the present study for the implementation of the systems based on fuzzy rules to predict the selection priorities of learners. Furthermore, a defuzzification method has been employed to supply crisp values from the fuzzy inputs, while Centroid of Area or COA was applied to defuzzify the results. The fuzzy inference system consisted of various MFs having five linguistic terms of Very Low, Low, Moderate, High, Very High (as shown in Fig. 11 and Fig. 12) in which adjustment of every MF has been done across the training phase. Discovery of fuzzy decision rules was possible using CART through forming neighborhoods in every cluster. It is worth noting that the number of fuzzy rules is determined according to the number of attributes found through LDA from the users' online reviews and the linguistic variables defined in the MFs.

The result for the prediction of user' overall interest in Udemy through fuzzy inferences system is shown in Fig. 13. It is found that there is a positive relationship between the quality factors (Knowledge of Instructor, Value of Information, Clarity of Explanation, Engaging Delivery of Information, Accuracy of Course Description, and Helpfulness of Practice Activities) and users' overall interests. This indicates that if the users receive high quality services from MOOCs websites, high level of satisfaction can be achieved which will impact on the intention to use MOOCs.

4.3. Evaluation metrics

This study has used several metrics including MAE, RMSE, precision, recall and F-measure for the method evaluation. They are explained in this section.

Confusion matrix (see Fig. 5) indicates the performance features employed in the calculation of AUC-ROC as well as AUC-PR where FN = False Negative, FP = False Positive, TP = True Positive, and TN = True Negative. AUC-ROC represents the area under the curve for the sensitivity (recall) plot against 1-specificity along thresholds. Sensitivity is according to all of the presences which have been observed, but specificity is according to all of the absences which have been observed or inferred. Therefore, all quadrants of the confusion matrix are used in AUC-ROC. The AUC-PR metric indicates the ROC of the plot of precision against recall or sensitivity along thresholds. Precision is according to all of the predicted presences. As a result, incorporation of the number of true negatives (TN) is not performed by AUC-PR. Here, precision indicates a measure of exactness, while recall indicates a measure of completeness. Despite accuracy and error rate, these two do not show sensitivity to modifications of data distributions. The F-measure metric indicated the "goodness" of a classifier when data imbalance is observed. A weighted harmonic mean of two metrics recall and precision forms F-measure as:

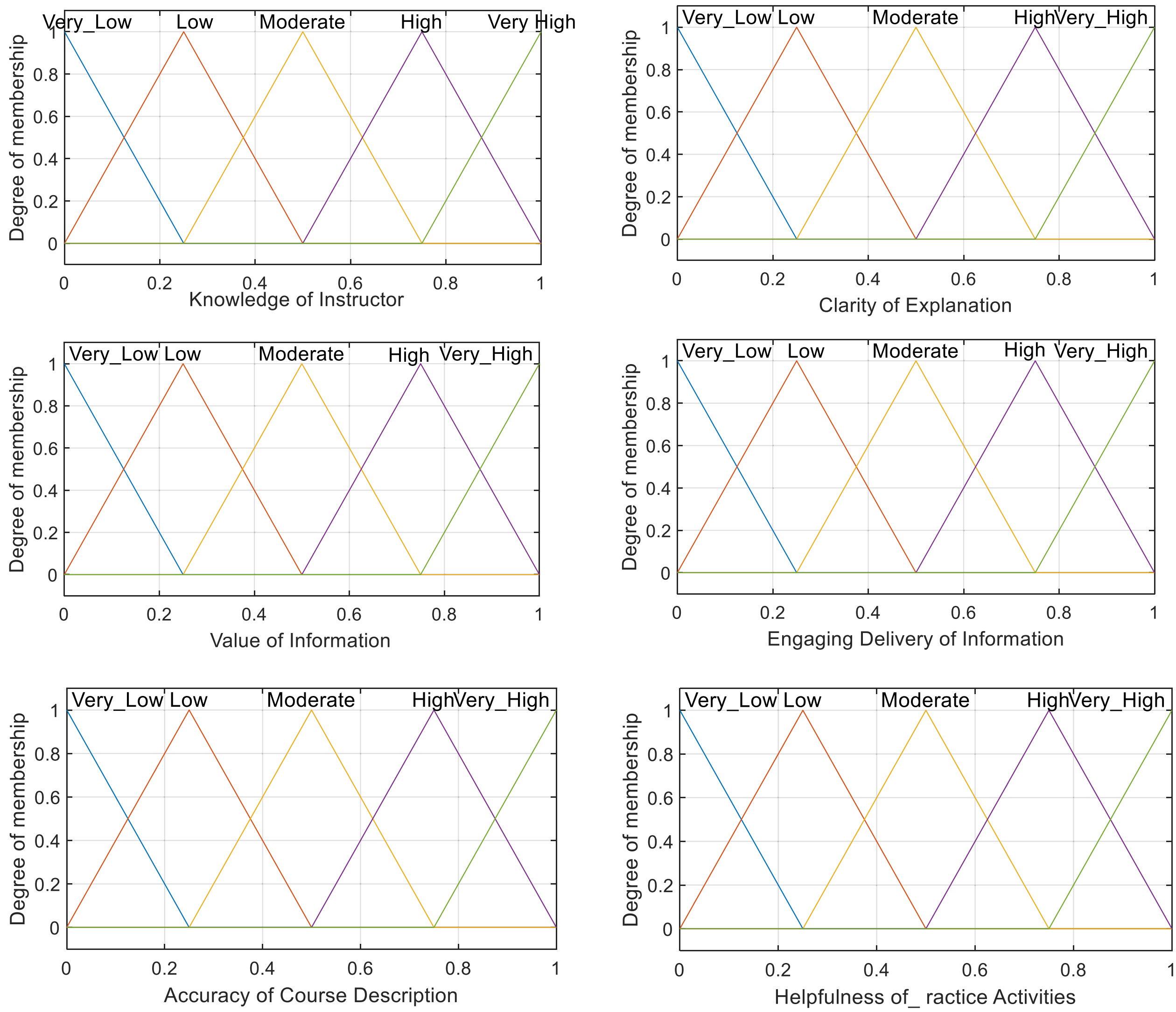


Fig. 11. A part of decision rules by CART in Cluster 1.

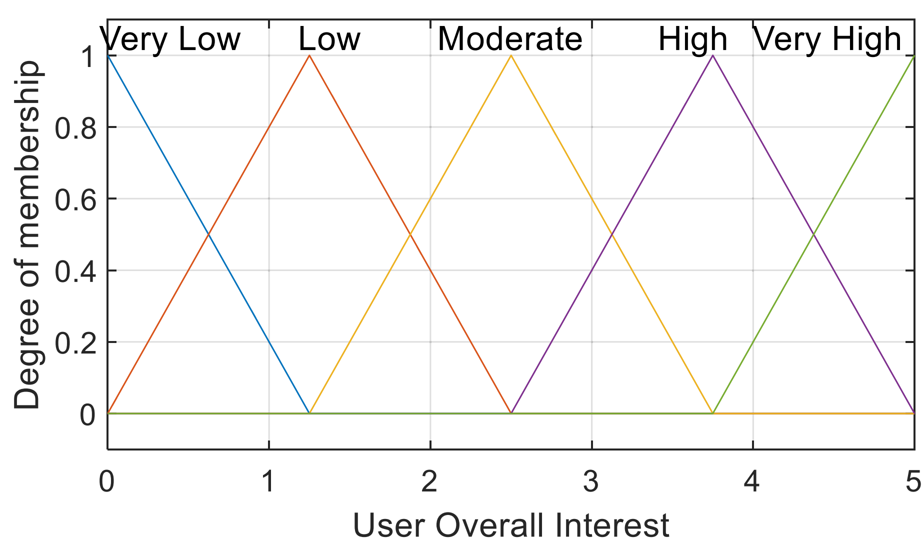


Fig. 12. A part of decision rules by CART in Cluster 1.

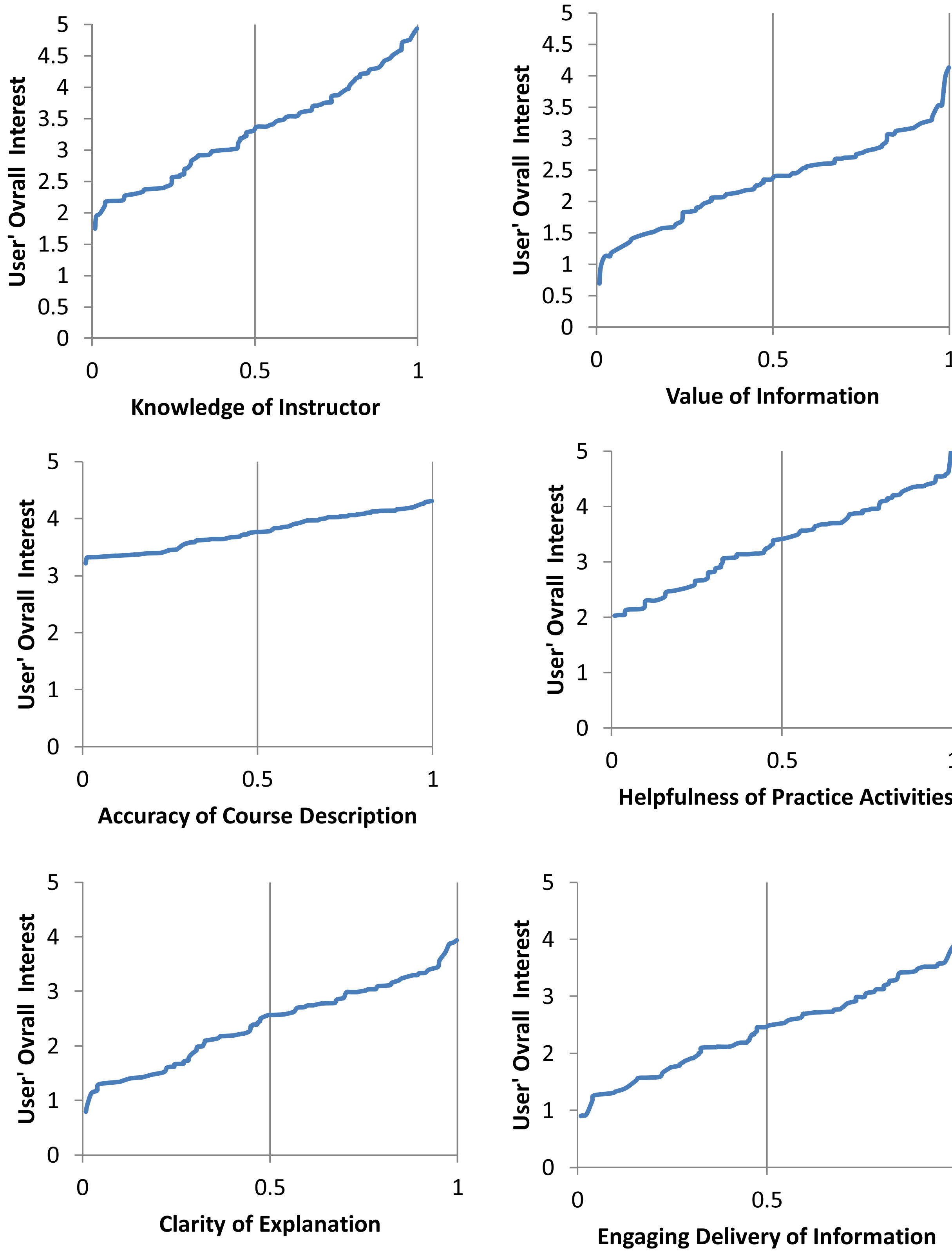


Fig. 13. Prediction of user' overall interest in Udemy through fuzzy inferences system.

Table 4
F1, RMSE and MAE results for the methods.

Methods	Random Selection			10-Fold Cross Validation		
	RMSE	MAE	F-measure	RMSE	MAE	F-measure
LDA + SOM + CART + Fuzzy Rule-Based	0.628	0.527	0.8342	0.584	0.514	0.8431
LDA + ANFIS	0.804	0.631	0.7778	0.738	0.621	0.7982
LDA + Fuzzy Rule-Based	0.811	0.672	0.7632	0.782	0.651	0.7751
SOM ensembles clustering	0.852	0.703	0.7332	0.826	0.682	0.7472
HOSVD	0.856	0.711	0.6153	0.833	0.687	0.6423
EuclideanCF	0.866	0.725	0.5323	0.852	0.704	0.5562
AvgSimCF	0.885	0.741	0.5146	0.868	0.724	0.5251
Standard CF	0.934	0.774	0.4927	0.891	0.752	0.5083

$$F = \frac{1}{\frac{\beta^2}{1+\beta^2} \times \frac{1}{Precision} + \frac{1}{1+\beta^2} \times \frac{1}{Recall}} = \frac{(1+\beta^2).p \times r}{\beta^2.Recall + Precision} \tag{19}$$

In which the weighting factor β denotes the corresponding significant of precision against recall. When $\beta = 1$, the relative score in F-measure is associated with the F1 score. In this research, we concentrate on the use of the F1-measure to evaluate the prediction accuracy in a multiclass classification model.

Note that, we used 10-fold cross validation and random selection approaches for method evaluation. The 10-fold cross validation approach has been simultaneously used for accurate performance assessment. It has been shown that this approach is superior to other methods (e.g., leave-one-out cross, holdout, and bootstrap validation methods) in model selection problems and determining the generalization error (Kohavi, 1995). The 10-fold cross validation approach splits the dataset into 10 equal or close-to-equal parts (folds). We also applied random selection of training and testing datasets for five times to obtain five training or testing groups in every cluster and subsequently come to higher accuracy of the evaluations. Division of the dataset into a training dataset which included 80% of users' ratings and a testing dataset which included rest of the ratings was carried out for each cluster. The results for these two approaches are presented and compared.

Mean Absolute Error (MAE) along with Root Mean Square Error (RMSE) were used for the evaluation of the predictive accuracy of the methods. MAE can be presented as follows:

$$MAE = \frac{\sum_{i,j \in T} |r_{ij} - \hat{r}_{ij}|}{|T|} \tag{20}$$

where r_{ij} indicates the rating by the user i given to item j, $\hat{r}_{ij}$ represents the predicted rating the user i given to item j. The number of tested ratings is presented by $|T|$. RMSE can be presented as follows:

$$RMSE = \sqrt{\frac{\sum_{i,j \in T} |r_{ij} - \hat{r}_{ij}|^2}{|T|}} \tag{21}$$

The adjusted coefficient of determination ($R^2_{adjusted}$) was used for the assessment of fuzzy-rule based models accuracy. The higher value of $R^2_{adjusted}$ and the lower MAE and RMSE indicate more desirable performance.

4.4. Multi-criteria recommendation performance

The effectiveness of multi-criteria recommendations compared to the single-rating CF technique has been previously proven in several studies. In this study, several tests were performed and the results were given for F1-measure (Onan & Toçoğlu, 2021), MAE and RMSE to provide a comparison of the proposed method with the methods introduced in the previous studies.

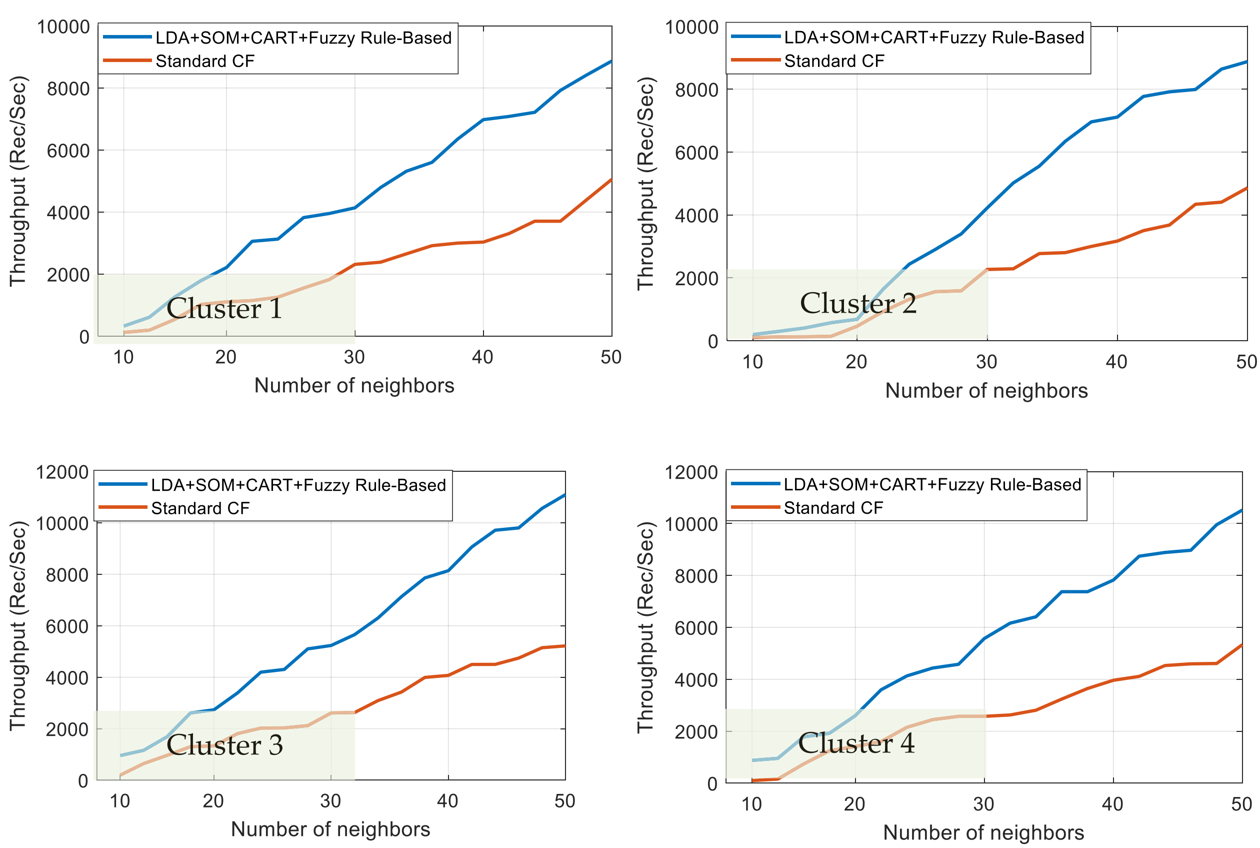


Fig. 14. Throughput of proposed method in four cluster of SOM.

Our method, LDA + SOM + CART + Fuzzy Rule-Based, is compared with the EuclideanCF (Liu, Mehandjiev, & Xu, 2011), AvgSimCF (Liu et al., 2011), Standard CF (Adomavicius & Kwon, 2007), HOSVD (Li, Wang, & Geng, 2008), SOM ensembles clustering (Tsai & Hung, 2012), LDA + ANFIS and LDA + Fuzzy Rule-Based. We used *k*-fold cross validation and random sampling for method evaluation. Both sampling approaches have been widely used in previous research (Agranoff et al., 2006; Mottaghitalab, Khanjari, Alizadeh, & Maghsoudi, 2021; Xiao, Ju, & Hui, 2017).

In cases of random selection and 10-fold cross validation approaches, the results for RMSE and MAE are shown in Table 4. The results showed that the 10-fold cross validation approach has provided better accuracies for all methods compared to the random selection approach. It is found that LDA + SOM + CART + Fuzzy Rule-Based provides the lowest RMSE and MAE followed by LDA + ANFIS, LDA + Fuzzy Rule-Based, SOM ensembles clustering, EuclideanCF, AvgSimCF, HOSVD and standard CF. In case of F1 evaluation results, it is seen that LDA + SOM + CART + Fuzzy Rule-Based (F1 = 0.8431) provides the best recommendations quality compared with the other recommendation methods. It is also found that SOM ensembles clustering method (F1 = 0.7472) provide relatively good results for recommendation quality over the methods such as EuclideanCF (F1 = 0.5562), AvgSimCF (F1 = 0.5251), Standard CF (F1 = 0.5083) and HOSVD (F1 = 0.6423). This method is based on single ratings. Two methods based on multi-criteria ratings, LDA + ANFIS (F1 = 0.7982) and LDA + Fuzzy Rule-Based (F1 = 0.7751), also outperform SOM ensembles clustering. This indicates that the use of multi-criteria rating can significantly influence the recommendation quality, as previous research found in the other contexts.

Evaluation of the suggested technique was performed for throughput which is the number of recommendations per second. It was sought to take various sizes of clustering into account using SOM to measure throughput. A comparison of the method with the Standard CF was carried out to indicate the throughput efficiency of the proposed method. The results are presented in Fig. 14. According to the plots, the throughput of the methods is indicated as a function of the number of neighbors. It can be obviously seen that the throughput of the suggested method that employs SOM, CART and Fuzzy Rule-Based techniques is higher than Standard CF method. Overall, it can be concluded that the proposed method is considerably capable of making the recommendation systems more scalable in the context of MOOCs.

5. Discussions

Providing online reviews in MOOCs websites which reveal the users' behaviour in online learning environments helps the learners select the most related courses according to their preferences. In such circumstances in which a significant population of learners and course events need considerable teaching materials, appropriate personalized intelligent systems will decrease the workload of the instructors who are engaged with helping the learners over the study period. Increasing the precision of MOOC websites in presenting the related information to the users can be considered as a significant activity which is of great importance in current educational data mining research. Multi-criteria CF recommending systems perform effectively in providing high accurate recommendations of courses to the users. However, there are a few studies focusing on the use of multi-criteria ratings in the CF system in educational systems.

The present study focused on the development of a recommendation agent for MOOCs based on multi-criteria CF and supervised as well as unsupervised machine learning methods. The proposed system has considered users' behaviour through the analysis of e-WOM in MOOCs. It was also aimed at accurate prediction of the learners' preferences in MOOCs according to multi-criteria ratings. One of the major contributions of the present study is in the generation of decision rules from the data with the use of decision trees. Analyzing the Udemy data indicated that the method performed effectively and found the suitable decision rules for educational materials in MOOCs. Moreover, the obtained decision rules were used in the proposed recommendation system based on fuzzy rules in order to help predict the overall learners' preference. Different tests performed on the real-world data set of Udemy have indicated the ability of proposed method to accurately predict the users' preference which could lead to significant improvement of the recommendation precision in MOOCs websites. Moreover, according to the obtained results of the Usemy data, the proposed technique is capable of performing well in the case of sparse data sets. Besides, clustering methods produced scalable recommendations of online resources in Udemy. Accordingly, the educational recommending systems can take advantage of these features and deal with problems of data sparsity, while taking assistance in the prediction of the courses with ratings from a limited population of users. Overall, given the effectiveness of the proposed method in improving the performance of educational recommendation systems, it has the capability of implementation in MOOCs websites to help users make decisions based on the high quality list of course recommendations.

6. Conclusion

The significant amount of educational material which is accessible across the Web necessitates existence of educational recommending systems which are able to effectively suggest the most suitable educational resources to the learners tailored to their preferences. The present paper proposed a recommendation agent to enrich web-based educational systems with accurate recommendations through multi-criteria ratings and considering their choice behaviour discovered from e-WOM. Data collection was performed using a crawler and generation of the significant criteria was carried out through LDA from Udemy data. SOM was employed to cluster the users' ratings regarding courses and fuzzy rule-based strategy was applied to predict the learners' preferences. Evaluation of the suggested recommendation methodology was performed on Udemy data with the use of RMSE, MAE, Precision as well as F1-measure. Evaluation results indicated that the proposed recommendation agent can predict accurately the learner' choice decisions when using LDA, CART, SOM and fuzzy rule-based methodology. This study has some limitations which can be addressed in the future studies. The inclusion of ensemble learning procedures to the recommendation methodology can be potentially investigated to further improve the accuracy of recommendation lists. Furthermore, the authors aim at extending the criteria from other resources including users' profiles, which would help in more accurate prediction of the users' preferences if taken into account. Another attractive area for future studies can be associated with the inclusion of dynamic feature learning methodologies to make incremental learning from the online reviews to provide more scalable recommendations.

CRediT authorship contribution statement

Mehrbakhsh Nilashi: Supervision, Conceptualization, Methodology, Investigation, Software, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Validation. **Behrouz Minaei-Bidgoli:** Conceptualization, Writing – review & editing, Visualization. **Abdullah Alghamdi:** Conceptualization, Methodology, Writing – review & editing, Validation. **Mesfer Alrizq:** Conceptualization, Writing – review & editing, Validation. **Omar Alghamdi:** Conceptualization, Writing – review & editing, Validation. **Fatima Khan Nayer:** Conceptualization, Writing – review & editing, Validation. **Nojood O Aljehane:** Conceptualization, Writing – review & editing, Validation. **Arash Khosravi:** Conceptualization, Writing – review & editing, Validation. **Saidatulakmal Mohd:** Conceptualization, Writing – review & editing, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Collaboration Funding program grant code NU/RC/SERC/11/10.

References

Abdelali, S. (2016). Education data mining: Mining MOOCs videos using metadata based approach. In *Paper presented at the 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)* (pp. 531–534).

Adomavicius, G., & Kwon, Y. (2007). New recommendation techniques for multicriteria rating systems. *IEEE Intelligent Systems, 22*(3), 48–55.

Agranoff, D., Fernandez-Reyes, D., Papadopoulos, M. C., Rojas, S. A., Herbster, M., Loosemore, A., et al. (2006). Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. *The Lancet, 368*(9540), 1012–1021.

Ahani, A., Nilashi, M., Yadegaridehkordi, E., Sanzogni, L., Tarik, A. R., Knox, K., et al. (2019). Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels. *Journal of Retailing and Consumer Services, 51*, 331–343.

Albahr, A., Che, D., & Albahar, M. (2021). A novel cluster-based approach for keyphrase extraction from MOOC video lectures. *Knowledge and Information Systems, 63*(7), 1663–1686.

Assami, S., Daoudi, N., & Ajhoun, R. (2018). Ontology-based modeling for a personalized MOOC recommender system. *Paper presented at the International Conference Europe Middle East & North Africa Information Systems and Technologies to Support Learning*, 21–28.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*(Jan), 993–1022.

Boroujeni, M. S., & Dillenbourg, P. (2019). Discovery and temporal analysis of MOOC study patterns. *Journal of Learning Analytics, 6*(1), 16–33.

Bousbahi, F., & Chorfi, H. (2015). MOOC-Rec: A case based recommender system for MOOCs. *Procedia-Social and Behavioral Sciences, 195*, 1813–1822.

Campos, R., dos Santos, R. P., & Oliveira, J. (2020). A Recommendation System based on Knowledge Gap Identification in MOOCs Ecosystems. *Paper presented at the XVI Brazilian Symposium on Information Systems*, 1–8.

Chakraborty, B., Chakma, K., & Mukherjee, A. (2016). A density-based clustering algorithm and experiments on student dataset with noises using Rough set theory. In *Paper presented at the 2016 IEEE International Conference on Engineering and Technology (ICETECH)* (pp. 431–436).

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357.

Cornelis, C., Lu, J., Guo, X., & Zhang, G. (2007). One-and-only item recommendation with fuzzy logic techniques. *Information Sciences, 177*(22), 4906–4921.

de Barba, P. G., Malekian, D., Oliveira, E. A., Bailey, J., Ryan, T., & Kennedy, G. (2020). The importance and meaning of session behaviour in a MOOC. *Computers & Education, 146*, Article 103772.

Fu, D., Liu, Q., Zhang, S., & Wang, J. (2015). The undergraduate-oriented framework of MOOCs recommender system. In *Paper presented at the 2015 International Symposium on Educational Technology (ISET)* (pp. 115–119).

Gunga, S. O., & Ricketts, I. W. (2008). The Prospects for E-Learning Revolution in Education: A philosophical analysis. *Educational Philosophy and Theory, 40*(2), 294–314.

Guo, S., Liu, Y., Chen, R., Sun, X., & Wang, X. (2019). Improved SMOTE algorithm to deal with imbalanced activity classes in smart homes. *Neural Processing Letters, 50*(2), 1503–1526.

Ha, S. H., Bae, S. M., and Park, S. C. (2000). ***Web mining for distance education.*** Paper presented at the Proceedings of the 2000 IEEE International Conference on Management of Innovation and Technology. ICMIT 2000.'Management in the 21st Century'(Cat. No. 00EX457), 715-719.

Hara, N., & Kling, R. (1999). Students' frustrations with a web-based distance education course. *First Monday, 4*(12).

Huang, R., & Lu, R. (2018). Research on Content-based MOOC Recommender Model. In *Paper presented at the 2018 5th International Conference on Systems and Informatics (ICSAI)* (pp. 676–681).

Hung, J.-L., & Crooks, S. M. (2009). Examining online learning patterns with data mining techniques in peer-moderated and teacher-moderated courses. *Journal of Educational Computing Research, 40*(2), 183–210.

Hwang, W.-Y., Chang, C.-B., & Chen, G.-J. (2004). The relationship of learning traits, motivation and performance-learning response dynamics. *Computers & Education, 42*(3), 267–287.

Jannach, D., Karakaya, Z., & Gedikli, F. (2012). Accuracy improvements for multi-criteria recommender systems. In *Paper presented at the Proceedings of the 13th ACM conference on electronic commerce* (pp. 674–689).

Kabra, R., & Bichkar, R. (2011). Performance prediction of engineering students using decision trees. *International Journal of computer applications, 36*(11), 8–12.

Khalid, A., Lundqvist, K., & Yates, A. (2020). Recommender systems for moocs: A systematic literature survey (january 1, 2012–july 12, 2019). *The International Review of Research in Open and Distributed Learning, 21*(4), 255–291.

Khalid, A., Lundqvist, K., Yates, A., & Ghzanfar, M. A. (2021). Novel online Recommendation algorithm for Massive Open Online Courses (NoR-MOOCs). *PLoS One, 16*(1), Article e0245485.

Khalil, M., & Ebner, M. (2017). Clustering patterns of engagement in Massive Open Online Courses (MOOCs): The use of learning analytics to reveal student categories. *Journal of Computing in Higher Education, 29*(1), 114–132.

Kim, K.-J., Liu, S., & Bonk, C. J. (2005). Online MBA students' perceptions of online learning: Benefits, challenges, and suggestions. *The Internet and Higher Education, 8*(4), 335–344.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Paper presented at the Ijcai*, 1137–1145.

Lal, T. N., Chapelle, O., Weston, J., & Elisseeff, A. (2006). Embedded methods. In *Feature extraction* (pp. 137–165). Springer.

Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., & Riera, T. (2014). A system for knowledge discovery in e-learning environments within the European Higher Education Area-Application to student data from Open University of Madrid, UDIMA. *Computers & Education, 72*, 23–36.

Li, Q., Wang, C., and Geng, G. (2008). ***Improving personalized services in mobile commerce by a novel multicriteria rating approach.*** Paper presented at the Proceedings of the 17th international conference on World Wide Web, 1235-1236.

Li, X., Li, X., Tang, J., Wang, T., Zhang, Y., & Chen, H. (2020). Improving Deep Item-Based Collaborative Filtering with Bayesian Personalized Ranking for MOOC Course Recommendation. In *Paper presented at the International Conference on Knowledge Science, Engineering and Management* (pp. 247–258).

Liang, J., Yang, J., Wu, Y., Li, C., & Zheng, L. (2016). Big data application in education: Dropout prediction in edx MOOCs. In *Paper presented at the 2016 IEEE Second International Conference on Multimedia Big Data (BigMM)* (pp. 440–443).

Liao, J., Tang, J., & Zhao, X. (2019). Course drop-out prediction on MOOC platform via clustering and tensor completion. *Tsinghua Science and Technology, 24*(4), 412–422.

Lin, Y., Feng, S., Lin, F., Zeng, W., Liu, Y., & Wu, P. (2021). Adaptive course recommendation in MOOCs. *Knowledge-Based Systems, 224*, Article 107085.

Liu, L., Mehandjiev, N., & Xu, D.-L. (2011). Multi-criteria service recommendation based on user criteria preferences. In *Paper presented at the Proceedings of the fifth ACM conference on Recommender systems* (pp. 77–84).

Mottaghitalab, A., Khanjari, A., Alizadeh, R., & Maghsoudi, H. (2021). Prediction of affinity coefficient for estimation of VOC adsorption on activated carbon using V-matrix regression method. *Adsorption*, 1–16.

Mukala, P., Buijs, J., & Van Der Aalst, W. (2015). Exploring students' learning behaviour in moocs using process mining techniques. *BPM Reports*, 1510.

Nilashi, M., Bin Ibrahim, O., & Ithnin, N. (2014). Hybrid recommendation approaches for multi-criteria collaborative filtering. *Expert Systems with Applications, 41*(8), 3879–3900.

Nilashi, M., Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. *Telematics and Informatics, 34*(4), 133–144.

Nilashi, M., Ibrahim, O., Samad, S., Ahmadi, H., Shahmoradi, L., & Akbari, E. (2019). An analytical method for measuring the Parkinson's disease progression: A case on a Parkinson's telemonitoring dataset. *Measurement, 136*, 545–557.

Nilashi, M., Jannach, D., Bin Ibrahim, O., & Ithnin, N. (2015). Clustering-and regression-based multi-criteria collaborative filtering with incremental updates. *Information Sciences, 293*, 235–250.

Nilashi, M., Yadegaridehkordi, E., Ibrahim, O., Samad, S., Ahani, A., & Sanzogni, L. (2019). Analysis of travellers' online reviews in social networking sites using fuzzy logic approach. *International Journal of Fuzzy Systems, 21*(5), 1367–1378.

Onan, A. (2018). An ensemble scheme based on language function analysis and feature engineering for text genre classification. *Journal of Information Science, 44*(1), 28–47.

Onan, A. (2019). Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering. *IEEE Access, 7*, 145614–145633.

Onan, A. (2020a). Mining opinions from instructor evaluation reviews: A deep learning approach. *Computer Applications in Engineering Education, 28*(1), 117–138.

Onan, A. (2020b). Sentiment Analysis in Turkish Based on Weighted Word Embeddings. In *Paper presented at the 2020 28th Signal Processing and Communications Applications Conference (SIU)* (pp. 1–4).

Onan, A. (2020). *Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach*. Computer Applications in Engineering Education.

Onan, A. (2020c). *Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks* (p. e5909). Concurrency and Computation: Practice and Experience.

Onan, A., Korukoğlu, S., & Bulut, H. (2016). Ensemble of keyword extraction methods and classifiers in text classification. *Expert Systems with Applications, 57*, 232–247.

Onan, A., Korukoğlu, S., & Bulut, H. (2017). A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. *Information Processing & Management, 53*(4), 814–833.

Onan, A., & Toçoğlu, M. A. (2020). Weighted word embeddings and clustering-based identification of question topics in MOOC discussion forum posts. *Computer Applications in Engineering Education*.

Onan, A., & Toçoğlu, M. A. (2021). A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification. *IEEE Access, 9*, 7701–7722.

Ouertani, H. C., & Alawadh, M. M. (2017). In *MOOCs recommender system: A recommender system for the massive open online courses* (pp. 139–143). Springer.

Peral, J., Maté, A., & Marco, M. (2017). Application of data mining techniques to identify relevant key performance indicators. *Computer Standards & Interfaces, 54*, 76–85.
Prabhakar, S., Spanakis, G., & Zaiane, O. (2017). Reciprocal recommender system for learners in massive open online courses (MOOCs). *Paper presented at the International Conference on Web-Based Learning*, 157–167.
Sakamoto, T. (2002). E-learning and educational innovation in higher education in Japan. *Educational Media International, 39*(1), 9–16.
Sebbaq, H., el Faddouli, N.-E., & Bennani, S. (2020). Recommender System to Support MOOCs Teachers: Framework based on Ontology and Linked Data. In *Paper presented at the Proceedings of the 13th International Conference on Intelligent Systems: Theories and Applications* (pp. 1–7).
Sheard, J., Ceddia, J., Hurst, J., & Tuovinen, J. (2003). Inferring student learning behaviour from website interactions: A usage analysis. *Education and Information technologies, 8*(3), 245–266.
Sloane, A. (1997). Learning with the Web: Experience of using the World Wide Web in a learning environment. *Computers & Education, 28*(4), 207–212.
Smart, K. L., & Cappel, J. J. (2006). Students' perceptions of online learning: A comparative study. *Journal of Information Technology Education: Research, 5*(1), 201–219.
Stošić, L., & Stošić, I. (2015). Perceptions of teachers regarding the implementation of the internet in education. *Computers in Human Behavior, 53*, 462–468.
Tang, H., Xing, W., & Pei, B. (2018). Exploring the temporal dimension of forum participation in MOOCs. *Distance Education, 39*(3), 353–372.
Tang, T. Y., & McCalla, G. (2002). Student modeling for a web-based learning environment: A data mining approach. *Paper presented at the AAAI/IAAI*, 967–968.
Tsai, C.-F., & Hung, C. (2012). Cluster ensembles in collaborative filtering recommendation. *Applied Soft Computing, 12*(4), 1417–1425.
Van den Beemt, A., Buijs, J., & Van der Aalst, W. (2018). Analysing structured learning behaviour in massive open online courses (MOOCs): An approach based on process mining and clustering. *International Review of Research in Open and Distributed Learning, 19*(5).
Wang, J., Xie, H., Au, O. T. S., Zou, D., & Wang, F. L. (2020). Attention-based CNN for personalized course recommendations for MOOC learners. In *Paper presented at the 2020 International Symposium on Educational Technology (ISET)* (pp. 180–184).
Wang, J., Xie, H., Wang, F. L., Lee, L.-K., & Au, O. T. S. (2021). Top-n personalized recommendation with graph neural networks in MOOCs. *Computers and Education: Artificial Intelligence, 2*, Article 100010.
Wang, P., Sanín, C., & Szczerbicki, E. (2013). Prediction based on integration of decisional DNA and a feature selection algorithm RELIEF-F. *Cybernetics and Systems, 44*(2–3), 173–183.
Xia, T. (2019). An E-learning Support Middleware with MOOC course recommendation. In *Paper presented at the 2019 14th International Conference on Computer Science & Education (ICCSE)* (pp. 596–600).
Xiao, X., Ju, W., and Hui, M. (2017). *iANOP-Enble: a sequence-based ensemble classifier for identifying antioxidant proteins by PseAAC and Random Forests.* Paper presented at the 2nd International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2017), 587-593.
Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior, 47*, 168–181.
Yousef, A. M. F., Chatti, M. A., Wosnitza, M., & Schroeder, U. (2015). A cluster analysis of MOOC stakeholder perspectives. *International Journal of Educational Technology in Higher Education, 12*(1), 74–90.
Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338–353.
Zhang, J., Hao, B., Chen, B., Li, C., Chen, H., & Sun, J. (2019). Hierarchical reinforcement learning for course recommendation in MOOCs. In *Paper presented at the Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 435–442).
Zhang, Z., Lin, H., Liu, K., Wu, D., Zhang, G., & Lu, J. (2013). A hybrid fuzzy-based personalized recommender system for telecom products/services. *Information Sciences, 235*, 117–129.