A User-Item Collaborative Filtering System to Predict Online Learning Outcome

Dina Fitria Murad*1, Rosilah Hassan2, Wahiza Wahi3, Bambang Dwi Wijanarko4

1Information Systems Department, BINUS Online Learning, Bina Nusantara University, Jakarta, 11480, Indonesia
2Center for Cyber Security, Faculty of Information Science and Technology, University Kebangsaan Malaysia. Selangor, 45500, Malaysia
3Pusat Citra Universiti (Centre for Liberal Studies), Universiti Kebangsaan Malaysia. Selangor, 45500, Malaysia
4Computer Science Department, BINUS Online Learning, Bina Nusantara University, Jakarta, 11480, Indonesia

ABSTRACT

Education has seen the rapid development of online learning. Many researchers have conducted studies on the use of recommendation systems in online learning. However, until now, several similar studies still focus on the accuracy of the prediction results. Various obstacles were encountered related to changes in the face to face learning process into online learning. This study uses the User-Item Collaborative Filtering method to predict student learning outcomes as a basis for providing recommendations to students. Data on student online learning outcomes were extracted using several methods as a basis for determining and assessing their learning outcomes. The dataset we use is dummy to match the original data. The findings of this study reveal that one of the reinforcing factors that affect student achievement in online learning is the quiz score. The students’ high achievement in the quizzes was also balanced by their active involvement during the learning process. Based on the evaluation of the recommendation system, it is known that the gradient boosted tree model is the best model for predicting the final score of student online learning with an accuracy calculated using the highest correlation of 0.7 and the smallest absolute error of 13.0 and root mean square error of 17.9. Based on the results of the evaluation, this study provides recommendations in the form of material links and learning archives that are useful for students to be able to carry out independent learning.

Keywords:
Grade prediction
Recommendation system
User-item base collaborative filtering
Online Learning
Learning Materials

1. Introduction

The Fourth Industrial revolution (4IR) has brought about significant impact to the global public and private sectors. One sector that has been immensely affected by the 4IR is education. This is especially true given the evolving technological approaches to teaching and learning applied by many educators across different levels of education. For example, the 4IR has seen a crucial role of learners when carrying out their learning activities and a flexibility of the approach and ambiance of learning offered to the learners as seen in many online classes conducted by higher learning institutions. This is illustrated in Graphic 1 below. Along the same line, many researches have reported on effective interactions amongst learners who are largely involve in the use of simulation, advanced technologies, and online learning as compared to the conventional methods of learning through book references [1].

In addition, the Internet of Things (IoT) is deemed essential in the educational settings nowadays as it is the fundamental tools required for any online learning to take place. The IoT devices are particularly useful in the process of interchanging information through wireless or wired network connections [2].

Online learning is a learning model that has become increasingly in demand among various groups of learners. On the same note, most researchers are currently focusing on finding the best means and approaches of online learning that fit in the continuous developments of the 4IR such as the strategies utilized by students when conducting online discussion [3] the use of
The escalating popularity of e-learning has instigated various challenges for higher learning institutions to provide massive educational resources and the dire need to search for relevant learning references. Generally, in a university environment that runs the e-learning process, students are equipped with some necessary materials for learning and some additional references. These large amount of learning resources sometimes leads to excessive information received by the students. Hence, this has sparked researchers’ interest to utilize a recommendation system that focuses on how students obtain their learning references either by using their profile [7], [8], or behavior [9], [10], and style or preferences which came to be known as personalization.

Starting with a basic theory, related to social studies [11], the recommendation system in the field of Education continues to develop and adapt several methods that successfully used in the field of e-commerce. Some of these studies state that the recommendation system in the area of Education needs personalized [12][13]. In general, the recommendation system can be grouped into four groups, namely (1) collaborative filtering, (2) content-base, (3) hybrid and (4) context-sensitive [14].

In the Education field, the collaborative filtering method is observed to be used effectively in the recommendation system [15], [16] in line with the success of the content-based approach [17], [18] and recently the most widely used method is hybrid [19], [20] and context-aware/context-sensitive [21], [22]. However, some researchers also found that rating determination in the recommendation system in the field of Education stated as a cold start [24]. In this regard, this study adopts a cognitive theory that describes learning as a business process which involves mental activities occurring in humans as a result of an active interaction with their environment to obtain a change in the form of knowledge, understanding, behavior, skills, values and attitude that are related and important. This study uses data on students online learning outcomes as a basis to determine and rate their learning materials. Subsequently, this study provides recommendations on the links to learning materials and archives that are useful for the students to be able to conduct their self-directed learning. Specifically for this study, a User-Item Collaborative Filtering method of the recommendation system is employed to predict the outcomes of students online learning. The set of data deployed for the purpose of this study consists of (1) student profiles, (2) learning process results and (3) contextual information regarding students’ perceptions and confidence in online learning.

Collaborative filtering techniques are used in recommendation systems as one technique for personalization [23], [24]. User base collaborative filtering (UBCF) works by collecting feedback from users in the form of ratings for items in a given domain. In our previous study [25] it was found that user collaborative filtering was the most dominant method for producing predictive accuracy. In this research we use User-Item collaborative filtering to predict the final grades of students by taking into account the personal values of students assessed based on learning activity (on student grades in other subjects). Meanwhile, item base collaborative filtering is used to predict the student’s final grade based on the proximity of the course that students have taken with other courses which are then averaged against other students in the same course.

Therefore, the collaborative filtering user-item technique is used by comparing the two techniques, UBCF with IBCF to maximize the predicted results of the student’s final grade.

This study is guided by this research question, “Which machine learning model has the highest accuracy to predict the value of online learning students”?.

2. Research Method

Using the preliminary data set that was gathered at the initial stage of the study which comprises of the completeness of the attributes, the filter of the outlier values, and the standardization of the data used, a combination of python and sklearn was further deployed to process the data. The stages of the research process are shown in Figure 2.

This study uses User-Item Collaborative Filtering method as preliminary data to find out the pattern of online learning process applied by students from previous session to be recommended to targeted students in this study. The structure of the data set used based on the results of the 2007-2018 student learning process at
BINUS Online Learning is shown in Table 1. We have dummy data set so that it maintains the confidentiality of the original student data. Although officially we have received permission from the institution.

The data set incorporates several attributes that are grouped together into 3 main parts which are:

- **Student profile (id and name)**
- **The output of the learning process (att, fod, pas1, pas2, qiz1, qiz2)**
- **Contextual information (student perception and student confidence)**

![Figure 3: Data set the attribute for UBCF](https://learning.oreilly.com/library/view/information-systems/9781482223712/chapter-47.html)

Points (2) and (3) represent student learning activities using the user-item base collaborative filtering method (figure 3) with various contextual information as student learning styles (figure 4).

![Figure 4: Attribute Dataset for user-item base collaborative filtering + Multi Context Information](https://learning.oreilly.com/library/view/information-systems/9781482223712/chapter-47.html)

3. **Result and Discussion**

In reference to the data set used, the student profile was the first input collated from this study to predict students’ grades. The student profile was complemented by the results of the learning process and students’ contextual information. Using the UBCF method, the value prediction is performed. The predictive value used to provide recommendations using a rule base and produce recommendations in the form of learning material links.

Prediction of student grades was made using several techniques including:

- Generalized linier model
- Deep learning
- Decision tree
- Random forest
- Gradient boosted trees
- Support vector machine

In reference to the six models above, the predicted value of student scores was compared with real student values. The results are presented in a scatter diagram in Figure 5.
Based on the results of the comparison, an evaluation was conducted using root mean square error, and the results are shown in Figure 6.

The results of the evaluation on the six learning models indicated that the gradient boosted trees model scored the most significant correlation with the smallest error compared to other models. Gradient boosted trees have the smallest root mean square error of 17.74 (generalized liner model 19.83, deep learning 18.85, decision tree 19.75, random forest 19.36, and support vector machine 23.45). The results were then broken down into parts as seen in Figure 7.

To complement the results of previous study [26] on the dominant factors that affect student attendance, we continue this research by processing data to look for correlations between each attribute (see Figure 8).

Based on the results of our previous research [25], it is known that the dominant factor affecting student achievement or failure of certain learning materials is the active involvement of students in discussion forums, followed by student achievement in quizzes. The results of this study are a little similar to previous studies except that the quiz score is placed as a dominant factor while the active involvement of students in discussion forums is a supporting factor.

4. Conclusion and Next Future

The results of this study complement our previous research regarding the attributes that most dominantly influence student achievement or failure, namely active participation in discussions / forums and quiz scores. These findings confirm that the active involvement of students in discussions / forums and quiz factors are important factors that contribute to meaningful and effective learning. In line with the fundamental social theory [11] and the cognitive learning theory which states that a person can learn from the success of others and tends to influence that person’s results, this study also proves that the success of the user collaborative method in e-commerce has also been successfully applied in the field of education. This method was also successful in influencing student learning outcomes. In this study, Gradient boosted trees model is the most accurate machine learning model that can predict student grades.

In the forthcoming research, we will conduct trainings on the use of the recommendation system. The results of the trainings will be our foundation to build machine learning models for the recommendation system in the field of Education, more specifically for online learning.

Conflict of Interest

This paper is the result of joint research. We declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

This work is supported by Bina Nusantara University through the Office of Research and Technology Transfer, as part of the Bina Nusantara University (BINUS) International Research Grant with Universiti Kebangsaan Malaysia (UKM) entitled "New Technology in Education Towards Fourth Industrial Revolution: Case Study Indonesia & Malaysia "contract number: No.026 / VR.RTT / IV / 2020 and contract date: April 6, 2020.

References


www.astesj.com