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

4 authors:

Dina Fitria Murad
Binus University
115 PUBLICATIONS   **503** CITATIONS

SEE PROFILE

Wahiza Wahi
Universiti Kebangsaan Malaysia
42 PUBLICATIONS   **206** CITATIONS

SEE PROFILE

Rosilah Hassan
Universiti Kebangsaan Malaysia
195 PUBLICATIONS   **2,508** CITATIONS

SEE PROFILE

Bambang Dw
Binus University
37 PUBLICATIONS   **185** CITATIONS

SEE PROFILE

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

Dina Fitria Murad[*,1], Rosilah Hassan[2], Wahiza Wahi[3], Bambang Dwi Wijanarko[4]

[1]Information Systems Department, BINUS Online Learning, Bina Nusantara University, Jakarta, 11480, Indonesia

[2]Center for Cyber Security, Faculty of Information Science and Technology, University Kebangsaan Malaysia. Selangor, 45500, Malaysia

[3]Pusat Citra Universiti (Centre for Liberal Studiess), Universiti Kebangsaan Malaysia. Selangor, 45500, Malaysia

[4]Computer Science Department, BINUS Online Learning, Bina Nusantara University, Jakarta, 11480, Indonesia

A R T I C L E   I N F O

A B S T R A C T

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.

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

various interface to enhance and encourage online learning [4], [5], or the production of expressive web designs [6].

	First Industrial Revolution	Fourth Industrial Revolution
Learning Objectives	Mastery of basic skills and knowledge (e.g., reading, math)	Development of whole person across multiple intelligences (e.g., emotional, intellectual, social)
Role of Educator	Expert	Facilitator
Learner Experience	'Factory model' - Passive, structured, directed, en masse	'Custom model' - Active, self-directed, exploratory
Target Age	K-12	Lifelong learning
Expertise	"Teacher knows best"	"Anyone can teach"
Access	Physical classroom	Anytime, anywhere, any device

Figure 1: Development of learning in education from first to fourth industrial revolution

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.
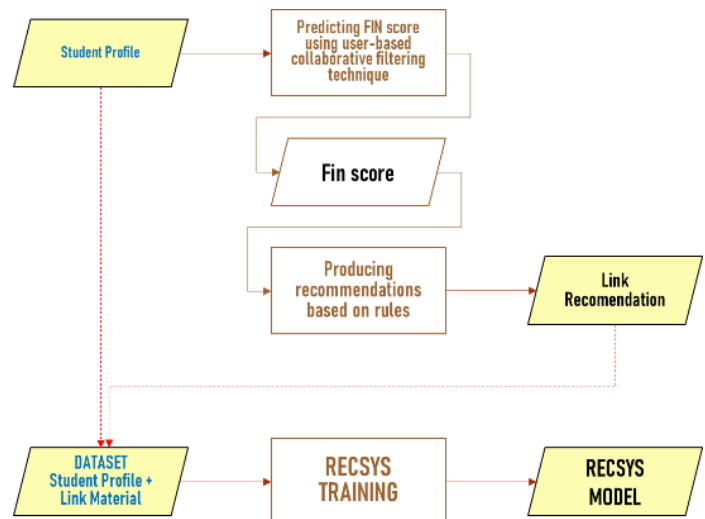

Figure 2: Research Framework

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 this 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)

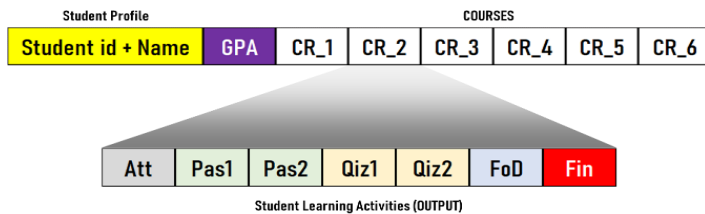- Contextual information (student perception and student confidence)

Figure 3: Data set the attribute for UBCF

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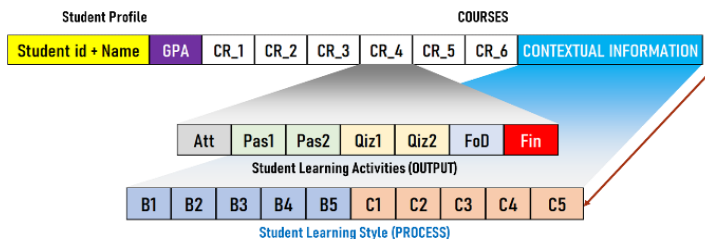

Figure 4: Attribute Dataset for user-item base collaboratife filtering + Multi Context Information

Table 1: Data Set Used

NAMA_MAHASISWA	ATT	FIN	FOD	PAS1	PAS2	QIZ1	QIZ2	TAS1	TAS2	TAS3	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	MATAKULIAH
1801437451_DONI	0	55	80	75	85	0	0	90	80	90	2	4	2	3	2	6	3	3	5	5	I0192_RESEAR
1801437451_DONI	0	49	60	100	100	0	0	82	50	100	2	4	2	3	2	6	3	3	5	5	K0442_QUAN
1801437451_DONI	0	82	0	0	0	0	0	0	0	0	2	4	2	3	2	6	3	3	5	5	M0244_INFOR
1801437451_DONI	0	94	60	97	95	0	0	100	90	90	2	4	2	3	2	6	3	3	5	5	M0304_CORP
1801437464_DANIE	0	82	100	100	100	0	0	100	100	100	2	4	3	4	5	5	4	5	4	4	CB432_CB: IN
1801437464_DANIE	0	55	100	85	80	0	0	80	80	90	2	4	3	4	5	5	4	5	4	4	I0192_RESEAR
1801437464_DANIE	0	25	100	100	60	0	0	80	65	100	2	4	3	4	5	5	4	5	4	4	K0442_QUAN
1801437464_DANIE	0	70	0	0	0	0	0	0	0	0	2	4	3	4	5	5	4	5	4	4	M0244_INFOR
1801437464_DANIE	0	94	100	80	97	0	0	99	100	90	2	4	3	4	5	5	4	5	4	4	M0304_CORP
1801437533_EKO R	0	86	100	100	100	0	0	100	100	100	2	5	3	3	4	5	5	5	4	5	CB432_CB: IN
1801437533_EKO R	0	60	95	75	80	0	0	90	80	60	2	5	3	3	4	5	5	5	4	5	I0192_RESEAR
1801437533_EKO R	0	39	100	40	100	0	0	95	65	100	2	5	3	3	4	5	5	5	4	5	K0442_QUAN
1801437533_EKO R	0	76	0	0	0	0	0	0	0	0	2	5	3	3	4	5	5	5	4	5	M0244_INFOR
1801437533_EKO R	0	100	90	97	95	0	0	100	90	90	2	5	3	3	4	5	5	5	4	5	M0304_CORP
1801437546_IMAM	0	76	100	100	100	0	0	100	100	100	2	2	4	4	4	4	4	4	4	3	CB432_CB: IN
1801437546_IMAM	0	90	100	80	80	0	0	75	70	70	2	2	4	4	4	4	4	4	4	3	I0192_RESEAR
1801437546_IMAM	0	72	100	40	100	0	0	80	65	100	2	2	4	4	4	4	4	4	4	3	K0442_QUAN
1801437546_IMAM	0	83	0	0	0	0	0	0	0	0	2	2	4	4	4	4	4	4	4	3	M0244_INFOR
1801437546_IMAM	0	94	90	0	95	0	0	99	100	90	2	2	4	4	4	4	4	4	4	3	M0304_CORP
1801437552_RAHM	0	78	90	100	100	0	0	100	100	100	2	6	4	4	3	3	3	6	5	3	CB432_CB: IN
1801437552_RAHM	0	55	70	75	80	0	0	70	80	80	2	6	4	4	3	3	3	6	5	3	I0192_RESEAR
1801437552_RAHM	0	70	75	40	100	0	0	50	60	100	2	6	4	4	3	3	3	6	5	3	K0442_QUAN
1801437552_RAHM	0	95	0	0	0	0	0	0	0	0	2	6	4	4	3	3	3	6	5	3	M0244_INFOR
1801437552_RAHM	0	89	65	100	95	0	0	90	100	90	2	6	4	4	3	3	3	6	5	3	M0304_CORP
1801437565_AGUN	0	76	50	85	88	0	0	97	90	97	2	3	4	4	4	4	4	4	4	3	CB432_CB: IN
1801437565_AGUN	0	55	55	84	94	0	0	76	83	70	2	3	4	4	4	4	4	4	4	3	I0192_RESEAR

with,

ATT	= attendance
FOD	= forum discussion
PAS1	= personal assignment1

PAS2 = personal assignment2
Qiz1 = quiz1
Qiz2 = quiz2
FIN = predictive target
B1-B5 = measure student perception toward courses
C1-C5 = measure student confidence toward courses
B1-B5, C1-C5 are multi-context information

In addition, a table that contains the list of materials (Refer to Table 2) is utilized to identify similar items related to student learning process and student context to be recommended to the targeted students. The list of materials provides a basis for the recommendations of students online learning outcomes on each learning material per topic per week. Learning outcomes are adjusted to the assessment of each learning material listed in the course outline with close reference to the Bloom taxonomy in line with the guidelines for course provisions as required by the higher learning institution in which the students are studying.

Table 2: List of Materials

M0124_ADVANCED INFORMATION S	10.001	LO1-The Systems Analyst and Information	https://www.or
ISYS6299_INFORMATION SYSTEMS (	10.001	LO1-Introduction to Information Systems	https://learning
M0124_ADVANCED INFORMATION S	10.002	LO1-Requirements Determination	https://www.or
ISYS6299_INFORMATION SYSTEMS (	10.002	LO1-Ethics;Privacy and Information Securit	https://learning
M0124_ADVANCED INFORMATION S	10.003	LO1-Use Case Analysis	https://www.or
ISYS6299_INFORMATION SYSTEMS (	10.003	LO1-Telecommunications; Networking;Wir	https://learning
M0124_ADVANCED INFORMATION S	10.004	LO2-Process Modeling	https://learning
ISYS6299_INFORMATION SYSTEMS (	10.004	LO2-Hardware and Software	https://learning
M0124_ADVANCED INFORMATION S	10.005	LO2-Data Modeling	https://learning
ISYS6299_INFORMATION SYSTEMS (	10.005	LO2-E-Business And E-Commerce	https://learning
M0124_ADVANCED INFORMATION S	10.006	LO2-The Design Phase;Design Strategy;and	https://learning
ISYS6299_INFORMATION SYSTEMS (	10.006	LO2-Social Computing	https://learning
M0124_ADVANCED INFORMATION S	10.007	LO2-User Interface Design	https://learning
ISYS6299_INFORMATION SYSTEMS (	10.007	LO2-Cloud Computing	https://learning
M0124_ADVANCED INFORMATION S	10.008	LO3-Program Design	https://learning
ISYS6299_INFORMATION SYSTEMS (	10.008	LO3-Customer Relationship Management a	https://learning
M0124_ADVANCED INFORMATION S	10.009	LO3-Data Storage Design	https://learning
ISYS6299_INFORMATION SYSTEMS (	10.009	LO3-Business Analytics	https://learning
M0124_ADVANCED INFORMATION S	10.01	LO3-Moving Into Implementation	https://learning
ISYS6299_INFORMATION SYSTEMS (	10.01	LO3-Acquiring Information Systems and Ap	https://learning
M0124_ADVANCED INFORMATION S	20.001	LO1-Requirements Determination	https://www.or

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.
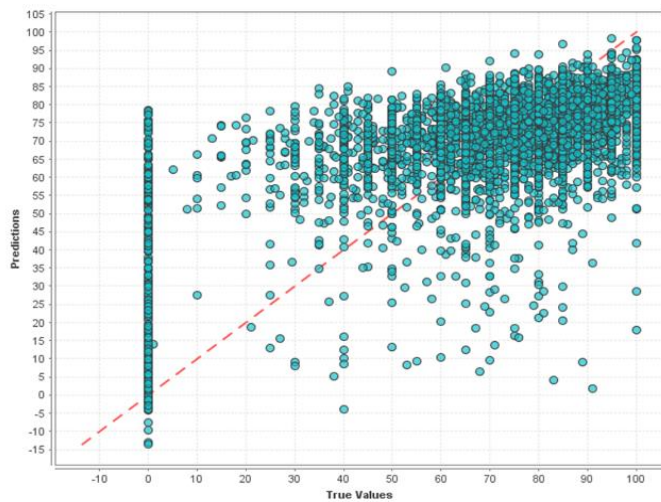
Figure 5: Scatter Diagram – Gradient Boosted Trees Model

Based on the results of the comparison, an evaluation was conducted using root mean square error, and the results are shown in Figure 6.

Model	Correlation	Standard Deviation	Squared Error	Standard Deviation	Relative Error	Standard Deviation	Absolute Error	Standard Deviation	Root Mean Squared Error	Standard Deviation
Generalized Linear Model	0.62	0.01	393.22	8.61	0.23	0.00	14.93	0.11	19.83	0.22
Deep Learning	0.67	0.02	355.41	15.02	0.23	0.00	13.95	0.24	18.85	0.40
Decision Tree	0.62	0.02	390.26	16.75	0.23	0.00	14.53	0.23	19.75	0.43
Random Forest	0.65	0.01	374.63	3.27	0.23	0.00	14.56	0.07	19.36	0.08
Gradient Boosted Trees	0.71	0.01	314.84	6.19	0.22	0.00	12.89	0.06	17.74	0.17
Support Vector Machine	0.47	0.01	550.42	30.97	0.22	0.01	15.99	0.53	23.45	0.67

Figure 6: Comparation: Error of Model

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 means square error of 17.74 (generalized linier 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.
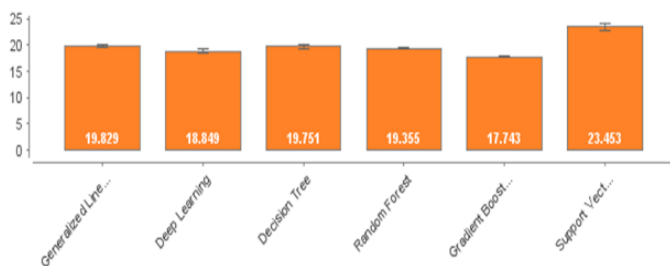

Figure 7: Comparation: Root Mean Square Error

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.

Figure 8: Correlation between Attribute

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 become 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

References

[1] Z. Mohd Tawil, N., Hassan, R., Ramlee, S., & K-Batcha, "Enhancing entrepreneurship skill among university's students by online business simulation," Journal of Engineering Science and Technology, 10 (Spec. Issue, (on 4th International Technical Conference (ITC)), 71–80, 2015.

[2] H. Mojib, G., Aman, A.H.M., Khalaf, M. y Hassan, "Simulation analysis for QoS in Internet of Things wireless network," 3C Tecnología. Glosas de Innovación Aplicadas a La Pyme, (Edición Especial), 77-83.

[3] A. Omar, Z. Amir, M. Mohamad, "Facilitating online learning: Students' online discussion strategies for a project work at a technical university in Malaysia," 3L: Language, Linguistics, Literature, **24**(4), 102–114, 2018, doi:10.17576/3L-2018-2404-08.

[4] A. Ramakrisnan, P., & Jaafar, "Usable, aesthetic, sociable and motivating interface for students' online knowledge sharing," In Learning and Collaboration Technologies : 3rd International Conference, LCT 2016 and Held as Part of HCI International (Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **9753**, 550–561, 2016.

[5] A. Ramakrisnan, P., & Jaafar, "Motivation design methodology for online knowledge sharing interface," In Advances in Visual Informatics - 5th International Visual Informatics Conference, IVIC 2017 (Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **10645 LNCS**, 224–232, 2017.

[6] Z. Daud, N. A., Redzuan, F., Nasruddin, Z. A., Sahari @ Ashaari, N., & Muda, "Persuasive web design for online islamic education," in In Proceedings of the 2017 6th International Conference on Electrical Engineering and Informatics: Sustainable Society Through Digital Institute of Electrical and Electronics Engineers Inc. Innovation, ICEEI 2017, 1–6, 2018.

[7] O. Bourkoukou, E. El Bachari, M. El Adnani, "A Personalized E-Learning Based on Recommender System," **2**(2), 99–103, 2016, doi:10.18178/ijlt.2.2.99-103.

[8] O. Khribi, M. K., Jemni, M., & Nasraoui, "Toward a hybrid recommender system for e-learning personalization based on web usage mining techniques and information retrieval," In E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, (pp. 6136-6145), 2007.

[9] N. Taghipour, A. Kardan, S. Ghidary, "Usage-based web recommendations: a reinforcement learning approach," … of the 2007 ACM Conference on …, 113–120, 2007, doi:10.1145/1297231.1297250.

[10] T. Mahmood, F. Ricci, "Improving recommender systems with adaptive conversational strategies," Proceedings of the 20th ACM Conference on Hypertext and Hypermedia - HT '09, (January), 73, 2009, doi:10.1145/1557914.1557930.

[11] A.J. Chournazidis, "Functionality and Feasibility of Knowledge Management in Enterprises," Procedia - Social and Behavioral Sciences, **73**, 327–336, 2013, doi:10.1016/j.sbspro.2013.02.059.

[12] N. Thai-Nghe, L. Drumond, T. Horváth, A. Nanopoulos, L. Schmidt-Thieme, "Matrix and Tensor Factorization for Predicting Student Performance," Undefined, 2011.

[13] A. Klašnja-Milićević, B. Vesin, M. Ivanović, Z. Budimac, "E-Learning personalization based on hybrid recommendation strategy and learning style identification," Computers and Education, **56**(3), 885–899, 2011, doi:10.1016/j.compedu.2010.11.001.

[14] B. Ricci, F., Rokach, L., & Shapira, Recommender systems: introduction and challenges, 2015.

[15] M. Dhanda, V. Verma, "Recommender System for Academic Literature with Incremental Dataset," Procedia Computer Science, **89**, 483–491, 2016, doi:10.1016/j.procs.2016.06.109.

[16] M.F. Caro Piñeres, J. Hernández, J.A. Jiménez Builes, "Diseño De Un Sistema De Recomendación En Repositorios De Objetos De Aprendizaje Basado En La Percepción Del Usuario: Caso Rodas," Ciencia e Ingeniería Neogranadina, **21**(1), 51–72, 2011.

[17] F. Matematika, P. Alam, U.G. Mada, "Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest," **4**(3), 194–200, 2017.

[18] F.O. Isinkaye, "Recommendation systems : Principles , methods and evaluation," Egyptian Informatics Journal, 2015, doi:10.1016/j.eij.2015.06.005.

[19] C. Wang, Z. Zheng, Z. Yang, "The research of recommendation system based on Hadoop cloud platform," in Proceedings of the 9th International Conference on Computer Science and Education, ICCCSE 2014, 193–196, 2014, doi:10.1109/ICCSE.2014.6926453.

[20] G. Article, "Recommender Systems - Comparison of Content-based Filtering and Collaborative Filtering," International Journal of Current Engineering and Technology, **4**(5), 3131–3133, 2014.

[21] T. Morrow, A.R. Hurson, S.S. Sarvestani, "A Multi-Stage Approach to Personalized Course Selection and Scheduling," 2017, doi:10.1109/IRI.2017.58.

[22] Y. Li, L. Mei, J. Wang, "A personalized recommendation system in E-Learning environment based on semantic analysis," in 2012 6th International Conference on New Trends in Information Science, Service Science and Data Mining (ISSDM2012), 2012.

[23] G. Adomavicius, A. Tuzhilin, "Context-aware recommender systems," Recommender Systems Handbook, Second Edition, 191–226, 2015, doi:10.1007/978-1-4899-7637-6_6.

[24] J.L. Herlocker, J.A. Konstan, L.G. Terveen, J.T. Riedl, "Evaluating collaborative filtering recommender systems," ACM Transactions on Information Systems, **22**(1), 5–53, 2004, doi:10.1145/963770.963772.

[25] D.F. Murad, Y. Heryadi, S.M. Isa, W. Budiharto, "Personalization of Study Material based on Predicted Final Grades using Multi-criteria User-collaborative Filtering Recommender System," Education and Information Technologies, **May**(III), 2020, doi:10.1007/s10639-020-10238-9.

[26] D.F. Murad, S.M. Isa, "Text Mining Analysis in the Log Discussion Forum for Online Learning Recommendation Systems," International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) 2018.