

Implementation of Learning Analytics in MOOC by Using Artificial Unintelligence

¹Budi Yulianto, ²Harjanto Prabowo, ¹Raymond Kosala and ¹Manik Hapsara

¹Department of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, Indonesia

²Department of Management, Bina Nusantara University, Jakarta, Indonesia 11480, Indonesia

Article history

Received: 23-11-2017

Revised: 18-12-2017

Accepted: 04-01-2018

Corresponding Author:

Budi Yulianto

Department of Computer

Science, Bina Nusantara

University, Jakarta, Indonesia

11480, Indonesia

Email: laboratory@binus.ac.id;

jibril_budi@yahoo.com

Abstract: Massive Open Online Course (MOOC), a web-based e-learning tool, is growing to be used by current educational institutions. To prevent high non-passing rate, instructor needs to know which learner has the potential to pass the course or not. Learner who will fail the course also need advices immediately from instructor or system to overcome it. Learning Analytics (LA) is needed to collect and analyze learners' activity logs on MOOC and predict their passing potential. The prototype application is developed by using Rational Unified Process (RUP) software development method. Implementation of LA in MOOC is feasible and suggested to analyze learners' success factors by consuming learners' activity logs and visualizing it in scatter diagram and node-link diagram. Instructor can provide advices to learners based on success factors generated by LA.

Keywords: Learning Analytics, Learner Success Factor, MOOC

Introduction

Massive Open Online Course (MOOC) is a web-based LMS that usually provides online and free (open) courses and can accommodate a significant (massive) number of learners. So, it allows learners to learn anywhere, anytime and on any device. MOOC is believed to increase learners' engagement and learning outcomes and it starts growing to be used by current educational institutions (Yulianto *et al.*, 2016a; 2017).

MOOC contains learning contents such as slide, video, audio, textbook, picture and also provide assessment to learner (Layona *et al.*, 2017). At the end of learning, learners will have exam to evaluate their learning outcomes. The evaluation result will be the reason to determine whether the learner will pass or fail the course.

Educational institutions expect every learner to pass all taken courses. To prevent high non-passing rate, instructor (or institution) needs to know which learner has the potential to pass the course or not. Learner who will fail the course needs advices immediately to overcome it (Yulianto *et al.*, 2016b).

MOOCs or other e-learning tools is suggested to implement Learning Analytics (LA) to collect and analyze learners' activity logs (records) and predict their passing potential. LA is the collection, measurement, analysis, reporting and prediction of data about learners and their contexts. LA is purposed

to understand and optimize learning and the environments in which it occurs (Siemens and Gasevic, 2012). LA involves predictive modeling and other advanced analytic techniques to generate learner's learning process and increase support needed to learner and instructor (Ruiz *et al.*, 2014).

Chatti implements LA by using B technique to detect hidden pattern by using education data set (Chatti *et al.*, 2012). There are 4 techniques such as statistics, information visualization, data mining and social network analysis.

LA is also implemented in some MOOCs such as SmartKlass, Blackboard Analytics and Open Academic Analytics Initiative (OAAI). Each method has different characteristics on cost, platform, or algorithm. SmartKlass provides machine learning algorithm. Blackboard Analytics gives web-based paid LA, can be used for general LMS and uses "black box learning" algorithm. Meanwhile, OAAI provides free LA but not for public, runs on Windows only, uses several machine learning algorithms such as decision tree, support vector machine and Bayesian Network.

Many data sets (factors) are used in LA to predict learning process. Erik conducted research of LA for online course with UNED COMA (Santos *et al.*, 2014). He stated that activities in forum is related with student passing rate. Activity in forum also shows individual ability according to Milligan (2015).

Warburton found that absence and duration to do assignment relate directly to learner’s final score. Meanwhile, seat position in class or sitting with specific group also affects positively to learner’s final score, where time difference between laboratory and theory does not affect (Akhtar *et al.*, 2017). Strang found that age, gender and culture did not affect directly to learner’s final score, but login times, reading and quiz activity did (Strang, 2017).

Based on previous researches conducted (Fig. 1), we can figure that grade (or final score) is commonly used by many systems as passing requirement (Layer 1). Some factors that influence grade directly are assignment, exam, or project score (Layer 2). Indirectly, factors that influence Layer 2 are activities recorded on systems technically (Layer 3) such as login time and count, post, submitted assignment, learning duration, etc. Last, factors that influence Layer 3 are hard to be recorded (quantified) such as age, gender, city, personality, learning style, previous skill, etc (Layer 4).

This study record and analyze data of Layer 3 (learning activities or logs) of learners whom pass and fail the course. After that, it gives recommendation to instructors and learners so they can take actions immediately to prevent failure (Yulianto *et al.*, 2013).

Many (machine/deep learning) algorithms or statistics methods are applied in LA, such as decision tree, support vector machine, Bayesian Network, back propagation, simple linear regression, paired t-test, etc. All are known as part of artificial intelligence methods. The term ‘intelligence’ means one’s capacity for logic, understanding, learning and problem solving. It can be stated as the ability to perceive information and to transform it as knowledge. Intelligence is commonly studied in humans, but currently has also been observed in inanimate such as machines or computers, known as ‘artificial intelligence’ (Albin, 2015).

Contradiction of artificial intelligence theory began to emerge. The results the machine analyzes, generates and predicts are the only data entirely rule-bound that machine itself doesn’t understand (Higgins, 1987; 1988). So, term ‘intelligence’ is not appropriately used. Paradoxically, if making humans more intelligent is not an easy job, how to do to machine? We should never assume that computers always get things right (Broussard, 2018). So, the term ‘unintelligence’ is proposed as contradiction to ‘intelligence’ for machine. This study uses some artificial unintelligence methods to analyze, generate and predict learner’s learning process.

Objectives of this study are using LA to consume recorded learners’ activity logs on MOOC, analyze and generate learner’s learning process. To visualize it, we implement a prototype application with scatter diagram to show students passing rate (Hai-Jew, 2015; Heymann and Le Grand, 2016) and node-link diagram by using simple linear regression and paired t-test statistics method to show learner learning success factors (Ward *et al.*, 2010; Illinois, 2016). At the end, LA is used to predict whether learners will pass or fail the course by showing the score predication in table, by using some artificial unintelligence methods such as back propagation, support vector machine, multiple linear regression and decision tree.

Method

Literature study is used to get common algorithms or methods used in previous studies for implementing LA. These algorithms will be used in prototype application to be implemented in MOOC. Prototype application development method in this study uses Rational Unified Process (RUP). This study will not discuss deeper about RUP since it’s not the objective and it can be easily explored online by readers.

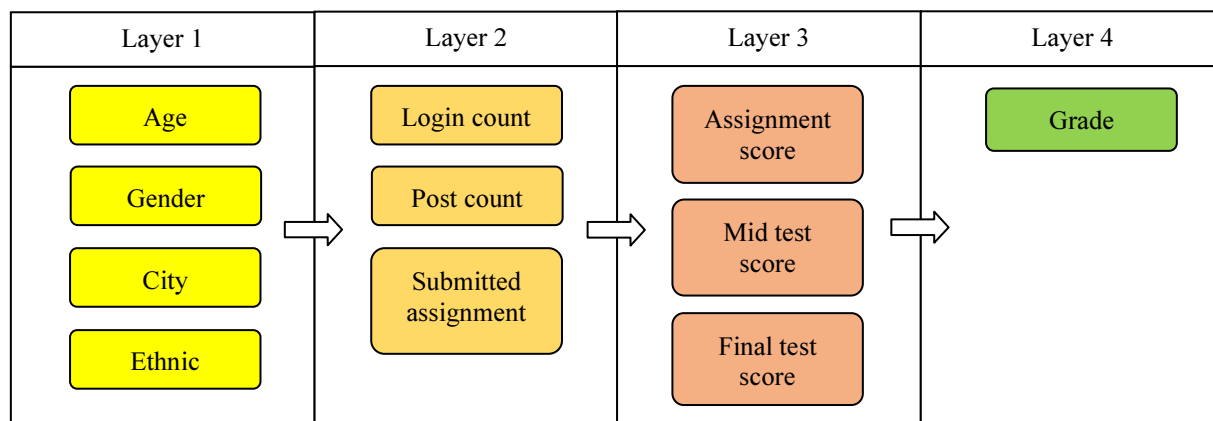


Fig. 1: Factors that influence learning outcomes

Prototype application uses statistics methods such as simple linear regression and paired t-test to consume and analyze recorded logs from learners' activities on MOOC and generate success factors. To predict learners' learning result (score), it uses artificial unintelligence algorithms such as back

propagation and support vector machine and statistics methods such as multiple linear regression and decision tree. The prototype application is developed by using Java and all the methods and algorithms are implemented by using JSAT library which is a library made by Edward Raff (2017a; 2017b).

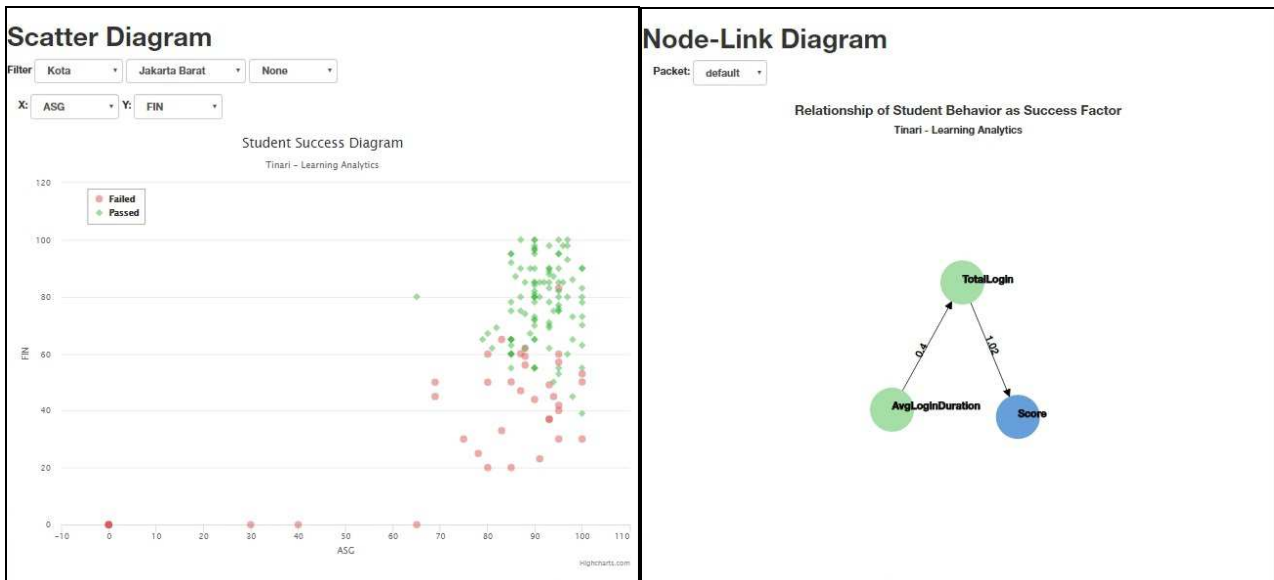


Fig. 2: Front-end application (Scatter and node-link diagram) view

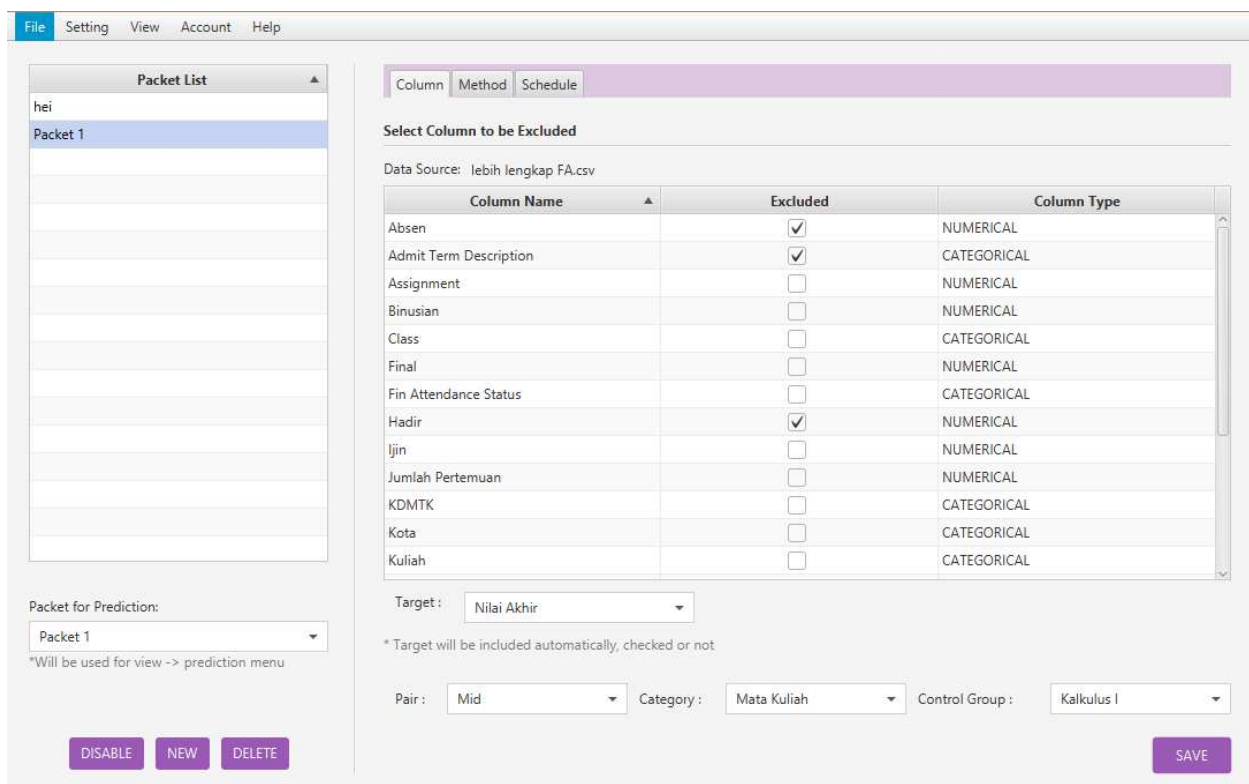


Fig. 3: Back-end application (database configuration) view

Proposed System

Prototype application is developed into two parts: front-end (web) and back-end (desktop). Front-end application is dedicated to guests, learners and instructors, while back-end is dedicated to administrator. Front-end application provides scatter diagram and node-link diagram (Fig. 2) and works well on various browsers such as Chrome, Firefox, Safari and IE and on various OS such as Windows, Linux and MacOS. It also developed in separate 3 web programming languages, which is PHP, JSP and C# ASP.NET. So, institutions can select appropriate one to be implemented in their existing MOOC system. Scatter diagram is used to show students passing rate and node-link diagram is used to show learner learning success factors.

Back-end application provides setting to be configured by administrator. Configuration includes database (Fig. 3) to be analyzed (server location,

tables and output), processing schedule (how many processes per day and on what day or date) and algorithm selection (for analyzing and predicting). It provides scatter and node-link diagram (Fig. 4) and prediction results (Fig. 5). Back-end application is developed on desktop-based and by using Java. It works well on Windows, Linux and MacOS.

Evaluation

Application was tested for prediction to know the accuracy of 4 methods. Evaluation is conducted by analyzing about 800 former learners' learning process logs (Layer 3). After that, equation models of each method are generated and tested to 20 new learners whom use internal institution MOOC system. Application will compare learners' actual and predicted score, then calculate the difference mean. The higher the difference has, the lower the accuracy is (Table 2).

Table 1: Results of evaluation

#	EV	BPNN		SVM		MLR		DT	
		PV	D	PV	D	PV	D	PV	D
1	66	63.27871	2.72129	68.30317	2.30317	62.30000	3.70000	65.6875	0.3125
2	61	52.15299	8.84701	63.62157	2.62157	57.50000	3.50000	60.8125	0.1875
3	68	64.29654	3.70346	71.16451	3.16451	66.57895	1.42105	68.2500	0.2500
4	68	69.26304	1.26304	72.26260	4.26260	66.69231	1.30769	69.1875	1.1875
5	72	69.26304	2.73696	72.26260	0.26260	66.69231	5.30769	69.1875	2.8125
6	76	77.90772	1.90772	75.25380	0.74620	73.66667	2.33333	72.1250	3.8750
7	83	82.99886	0.00114	84.22741	1.22741	78.40000	4.60000	80.8750	2.1250
8	66	65.31087	0.68913	67.50816	1.50816	61.20000	4.80000	64.6875	1.3125
9	49	47.80357	1.19643	54.17092	5.17092	50.06667	1.06667	51.9375	2.9375
10	41	46.42355	5.42355	55.23480	14.2348	46.60000	5.60000	52.5000	11.500
11	52	52.09498	0.09498	56.84095	4.84095	53.33333	1.33333	54.6875	2.6875
12	81	80.62458	0.37542	74.08013	6.91987	73.86667	7.13333	71.2500	9.7500
13	89	83.30288	5.69712	84.71859	4.28141	78.40000	10.6000	81.3750	7.6250
14	79	83.69095	4.69095	84.14316	5.14316	84.33333	5.33333	81.0625	2.0625
15	85	84.03742	0.96258	90.59394	5.59394	86.71429	1.71429	87.1250	2.1250
16	89	84.13513	4.86487	91.43777	2.43777	86.71429	2.28571	88.0000	1.0000
17	86	83.86038	2.13962	84.63435	1.36565	84.33333	1.66667	81.5625	4.4375
18	89	84.35101	4.64899	88.84484	0.15516	86.46154	2.53846	86.0000	3.0000
19	85	83.08144	1.91856	86.82581	1.82581	86.46154	1.46154	83.4375	1.5625
20	85	81.29195	3.70805	82.75386	2.24614	78.40000	6.60000	79.3125	5.6875
Difference Mean		2.8795435		3.51559		3.7151545		3.321875	

EV: Expected Value (Actual Score); BPNN: Backpropagation; SVM: Support Vector Machine; MLR: Multiple Linear Regression; DT: Decision Tree; PV: Predictive Value (Score); D: Difference or |EV-PV|

Table 2: Results of Speed Performance (in ms)

#	Scatter Diagram				Node-Link Diagram			
	1	2	3	4	1	2	3	4
1	25.000	25.000	22	20	87	99	95	93
2	25.000	24.000	13	22	82	81	95	97
3	22.000	20.000	22	25	95	80	77	78
4	18.000	22.000	19	21	95	106	90	112
Mean	21.5625	91.375						

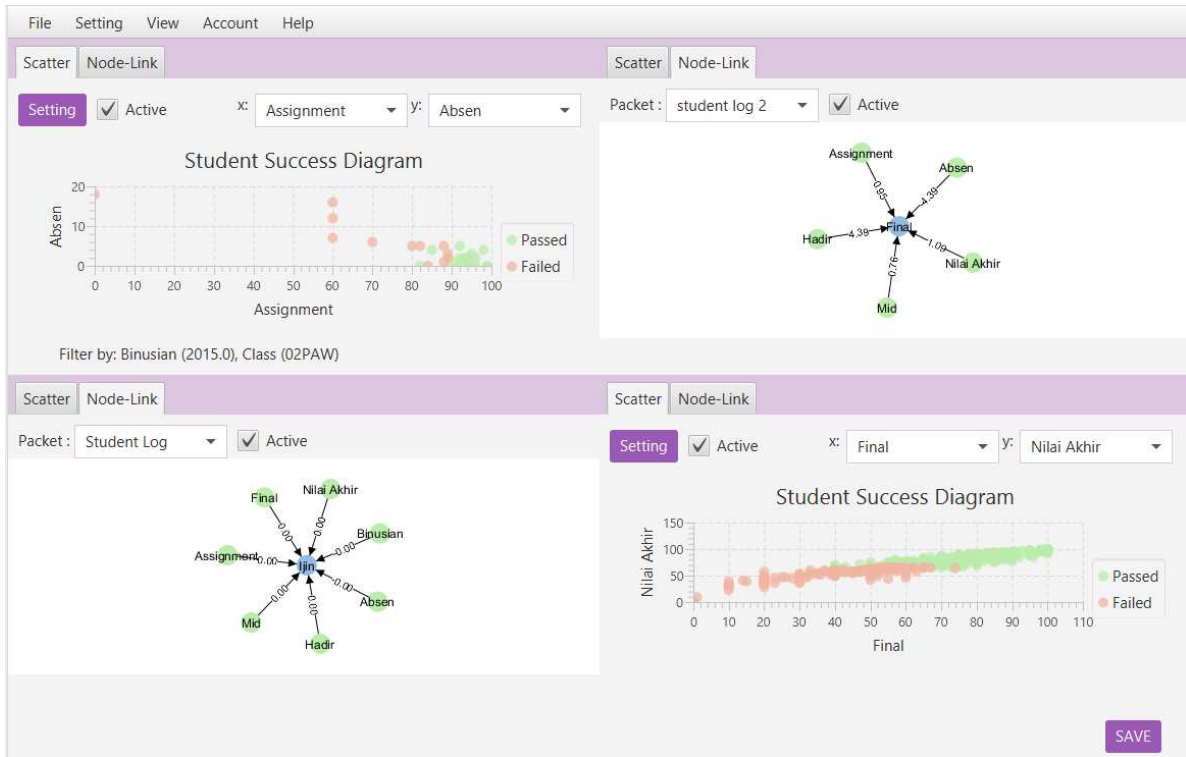


Fig. 4: Back-end application (scatter and node-link diagram) view

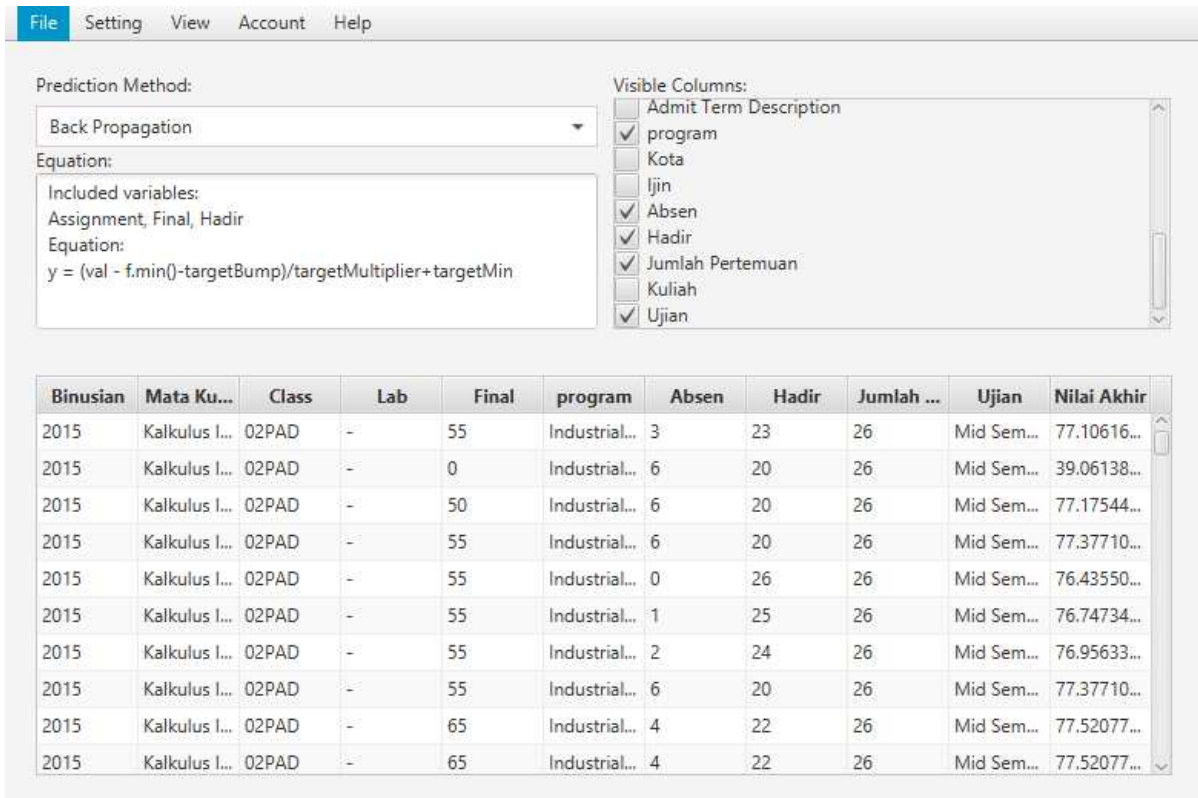


Fig. 5: Back-end application (prediction results) view

Accuracy result shows that prediction by using BPNN has difference mean of 2.88, which is smaller than other algorithms. But, it does not mean that BPNN gives better accuracy result for prediction. It can depend on many factors and institution is suggested to try other algorithms.

Evaluation is also conducted to test back-end speed performance for analyzing and visualizing report in scatter and node-link diagram. There are 500 data used in this evaluation and 16 experiments for each. It needs about 21.6 milliseconds to analyze and visualize scatter diagram and about 91.4 milliseconds for node-link diagram.

Conclusion

LA is feasible and suggested to be implemented in MOOC to collect and analyze learners' success factors by consuming learners' activity logs and visualizing it in scatter diagram and node-link diagram. By this prototype application, LA is implemented and success factors is modelled by using simple linear regression and paired t-test statistics method. To predict learners' score is by using back propagation, support vector machine, multiple linear regression and decision tree. Instructor and institutions are expected to prevent immediately if a learner is suspected to fail the course.

For further researches, it's suggested to add more methods or algorithms to analyze learners' success factors and predict algorithm for student success, so that admin has more options and the result can be compared with another analysis and prediction method. Based on success factors, MOOC can provide recommender system and adaptive system. Recommender system is needed to give recommendation to learner what should do and do not to pass the course. Learner is freed to follow or not the recommendation. Adaptive system is needed to adjust the system based on learner profile (Yulianto and Prabowo, 2017).

Acknowledgment

This research is funded by the Ministry of Research, Technology and Higher Education of the Republic of Indonesia through Doctoral Dissertation Grant (contract number of 039A/VR.RTT/VI/2017); and supported by Doctor of Computer Science, Bina Nusantara University. Thanks to Winarti Iswin, Riyn Winesdyo Winarko and Tiffany Inne for developing the prototype.

Authors Contribution

Budi Yulianto: Lead research project, coordinate developer, doing experiment, be an instructor, data analysis and writing the manuscript.

Harjanto Prabowo: Advise research project, design the experiment, data analysis and writing manuscript.

Raymond Kosala: Advise research project, design the application, data analysis, writing manuscript, proof reading.

Manik Hapsara: Advise research project, design the research methodology, data analysis, writing manuscript, proof reading.

Ethics

Authors confirm that this manuscript has not been published elsewhere and that no ethical issues are involved.

Conflict of Interest Declaration

Authors declare that there is no conflict of interest regarding the publication of this manuscript.

References

- Akhtar, S., S. Warburton and W. Xu, 2017. The use of an online learning and teaching system for monitoring computer aided design student participation and predicting student success. *Int. J. Technol. Design Educ.*, 27: 251-270. DOI: 10.1007/s10798-015-9346-8
- Albin, J., 2015. *Artificial Unintelligence: The First AI Tragedy*. 1st Edn., Content Arcade Publishing, ISBN-10: 1517758351, pp: 36.
- Broussard, M., 2018. *Artificial unintelligence: How computers misunderstand the world*. The MIT Press.
- Chatti, M.A., A.L. Dyckhoff, U. Schroeder and H. Thüs, 2012. A reference model for learning analytics. *Int. J. Technol. Enhanced Learn.*, 4: 318-331. DOI: 10.1504/IJTEL.2012.051815
- Hai-Jew, S., 2015. What is a network graph? What is a node-link diagram?
- Heymann, S. and B. Le Grand, 2016. *Graph Viz* 101.
- Higgins, J., 1987. *Artificial unintelligence: Computer uses in language learning*. *TESOL Quarterly*, 21: 159-165. DOI: 10.2307/3586364
- Higgins, J., 1988. *Language, learners and computers: Human intelligence and artificial unintelligence*. Addison-Wesley Longman Ltd.
- Illinois, 2016. *Scatter plot*.
- Layona, R., B. Yulianto and Y. Tunardi, 2017. Authoring tool for interactive video content for learning programming. *Proc. Comput. Sci.*, 116: 37-44. DOI: 10.1016/j.procs.2017.10.006
- Milligan, S., 2015. Crowd-sourced learning in MOOCs: Learning analytics meets measurement theory. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge*, Mar. 16-20, ACM, Poughkeepsie, New York, pp: 151-155. DOI: 10.1145/2723576.2723596

- Raff, E., 2017a. JSAT. <https://github.com/EdwardRaff/JSAT>
- Raff, E., 2017b. Project details for java statistical
- Ruiz, J.S., H.J.P. Díaz, J.A. Ruipérez-Valiente, P.J. Muñoz-Merino and C.D. Kloos, 2014. Towards the development of a learning analytics extension in open edX. Proceedings of the 2nd International Conference on Technological Ecosystems for Enhancing Multiculturality, Oct. 01-03, ACM, Salamanca, Spain, pp: 299-306. DOI: 10.1145/2669711.2669914
- Santos, J.L., J. Klerkx, E. Duval, D. Gago and L. Rodríguez, 2014. Success, activity and drop-outs in MOOCs an exploratory study on the UNED COMA courses. Proceedings of the 4th International Conference on Learning Analytics and Knowledge, Mar. 24-28, ACM, Indianapolis, Indiana, USA, pp: 98-102. DOI: 10.1145/2567574.2567627
- Siemens, G. and D. Gasevic, 2012. Guest editorial-learning and knowledge analytics. *Educ. Technol. Society*, 15: 1-2.
- Strang, K.D., 2017. Beyond engagement analytics: Which online mixed-data factors predict student learning outcomes? *Educ. Inform. Technol.*, 22: 917-937. DOI: 10.1007/s10639-016-9464-2
- Ward, M.O., G. Grinstein and D. Keim, 2010. *Interactive Data Visualization: Foundations, Techniques and Applications*. 1st Edn., CRC Press, ISBN-10: 143986554X, pp: 513.
- Yulianto, B. and H. Prabowo, 2017. Effective digital contents for computer programming learning: A systematic literature review. *Adv. Sci. Lett.*, 23: 4733-4737. DOI: 10.1166/asl.2017.8877
- Yulianto, B., E. Heriyanni, R.E. Sembiring, R. Amalia and R. Fridian, 2013. Aplikasi pembelajaran algoritma dasar interaktif berbasis computer assisted instruction. *ComTech: Comput. Math. Eng. Applic.*, 4: 1255-1266. DOI: 10.21512/comtech.v4i2.2611
- Yulianto, B., H. Prabowo, R. Kosala and M. Hapsara, 2016a. MOOC architecture model for computer programming courses. Proceedings of the International Conference on Information Management and Technology, IEEE Xplore Press, Nov. 16-18, IEEE Xplore Press, Bandung, Indonesia, pp: 35-40. DOI: 10.1109/ICIMTech.2016.7930298
- Yulianto, B., H. Prabowo and R. Kosala, 2016b. Comparing the effectiveness of digital contents for improving learning outcomes in computer programming for autodidact students. *J. e-Learn. Knowl. Society*.
- Yulianto, B., H. Prabowo, R. Kosala and M. Hapsara, 2017. Harmonik =++(Web IDE). *Proc. Comput. Sci.*, 116: 222-231. DOI: 10.1016/j.procs.2017.10.044