

Comparative Study on the Different Testing Techniques in Tree Classification for Detecting the Learning Motivation

C Juliane*, A A Arman, H S Sastramihardja, and I Supriana

School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia

^{1*}christina.juliane@students.itb.ac.id

^{2,3,4}[arman|husni|iping]@stei.ac.id

Abstract. Having motivation to learn is a successful requirement in a learning process, and needs to be maintained properly. This study aims to measure learning motivation, especially in the process of electronic learning (e-learning). Here, data mining approach was chosen as a research method. For the testing process, the accuracy comparative study on the different testing techniques was conducted, involving Cross Validation and Percentage Split. The best accuracy was generated by J48 algorithm with a percentage split technique reaching at 92.19 %. This study provided an overview on how to detect the presence of learning motivation in the context of e-learning. It is expected to be good contribution for education, and to warn the teachers for whom they have to provide motivation.

1. Introduction

Learning is a change in behavior as a result of the process repeated regularly in an environment [1]. The changes are in positive way, so it can be an indication that explains to what extent a person has successfully learned by achieving certain parameters. According to the direction of National Ministry of Education for Republic of Indonesia in 2008, there are some principles to be held in learning a management process. Learning should be interactive, inspiring, fun, challenging, and motivating.

The interactive, inspiring, fun, and challenging principles in learning will certainly build up the learning motivation. However, the level of learning motivation will affect the success rate of learning process. Motivation can be identified as a process carried out with some intensity and persistence values [2]. Intensity is defined as how hard a person's effort is, while persistence is how hard a person's can maintain his effort to achieve the goal. This study provided an overview how to detect the presence of learning motivation in the context of e-learning. It is expected to be good contribution for education, and to warn teachers to whom they have to provide motivation. Data mining approach with classification technique was used to identify the presence of learning motivation. Some of Tree Classifier algorithms and testing technique are compared and expected would generate a good prediction. Learning motivation prediction has been built by WEKA toolkit for Tree Classifier [3] to validate a total of 1920 numbers of instances and compared their performance using Cross Validation and Percentage Split techniques. The organization of the paper is as follows: Part 2 presents some relevant research works carried out in this field, Part 3 presents how the research method was conducted, and Part 4 and Part 5 explain how to build the classification and compare the performance by using the classification techniques. The final conclusion of our study is discussed in Part 6.



2. Related Rork

Detecting the presence of learning motivation has been carried out by several researchers and methods. Some researcher used a framework such as Deci & Ryan Framework of Motivation and NEO-FFI Personality Questionnaire to predict the presence of learning motivation [4][5]. But others used a variety of approaches in the field of computer science such as facial expression recognition [6] and data mining [7] to do the same thing. Using those frameworks and approaches, however, still could not identify the presence of learning motivation in the context of IT-based learning.

Using web content mining and web usage mining on the variable access time and relevance of teaching content could not identify the presence of learning motivation in the real time [7]. [8] have identified three groups of attribute to identify the learning motivation in real-time learning, those are velocity in the process of e-learning (v), quantity of the answer (q), and relevancy of the answers (r). This study was carried out to predict the learning motivation in the context of electronic learning by using the attribute of velocity, quantity, and relevancy [8]. Tree classifier was chosen to build the prediction of learning motivation and compared the performance by WEKA tree classifier toolkit and several testing techniques such as Cross Validation and Percentage Split. Considering the simplicity to interpret the result of prediction, Tree classier algorithm was then chosen [9]. A subsection

3. Research Method

Data preparation, data collection, data processing, and data testing were the stages of research methodology conducted in this research. Preparing a list of questions and selecting the potential respondents to the field of the study were carried out in the data preparation. Distributing the questionnaire to the respondent and gathering the result were conducted in the process of data collection. Data processing and data testing were carried out by using tree classifier in WEKA toolkit [3], tested and comparing the performance by cross validation and percentage split techniques. The research methodology can be seen in Figure 1.

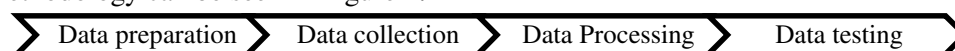


Figure 1. Research Methodology

4. Classification Techniques Sections, subsections and subsubsections

The classifier algorithm in WEKA toolkit such as AD Tree, BF Tree, Decision Stump, FT, ID3, J48, J48graft, LAD Tree, LMT, NB Tree, Random Forest, Random Tree, REP Tree, and Simple Cart was used to process the data training of 1920 of instances. The attributes as seen in Table 1 were used in the classification.

Table 1. Attributes to predict the learning motivation [8].

Group of Attributes	Symbol	Description	Value
Velocity	v_1	Speed of answering questions/assignments/quizzes/exams/email/chat	Fast-Slow
	v_2	Speed of downloading learning material/assignment/exams/quizzes question	Fast-Slow
	v_3	Speed of uploading assignments answer/ quizzes answer/ tests answer	Fast-Slow
Quantity	q	Quantity (a length) of the answer of assignment/exams/quizzes	A Bit-Many
Relevancy	r_1	The correctness of answering questions/assignment/exams/email	False-True
	r_2	Following the rules or standards	Neglect-Follow
Motivation	m	The value of learning motivation	Low-High

The training process was performed by dividing a total of 1920 numbers of instances into 3 sets of 640, 1280, and 1920 instances and compared the accuracy of tested algorithm. The accuracy used here consisted of time used in building the model (time/second), the value of correctly, incorrectly, precision, recall, f-measure, and ROC (Receiver Operating Characteristic) [10].

5. Classification Comparative

Cross Validation and Percentage Split techniques were conducted to compare the accuracy result. Ten-fold cross validation was selected for data testing, meaning that total 1920 of instances would be split into 10 parts and randomly formed by the principle of 1:9. It means that one part would be as data testing, and other nine were used as data training. This process was continued until all parts had an opportunity to be a data testing and accuracy was measured [9]. The accuracy would be compared by other testing techniques, which was percentage split. In percentage split, it means that 60% of instances would be used as data training, and 40% of instances would be used as the data testing. The result of these testing techniques was compared to observe the best accuracy performed. The best accuracy was conducted in the set of 1280 instances for the two testing techniques and it can be seen in Table 2 and Table 3.

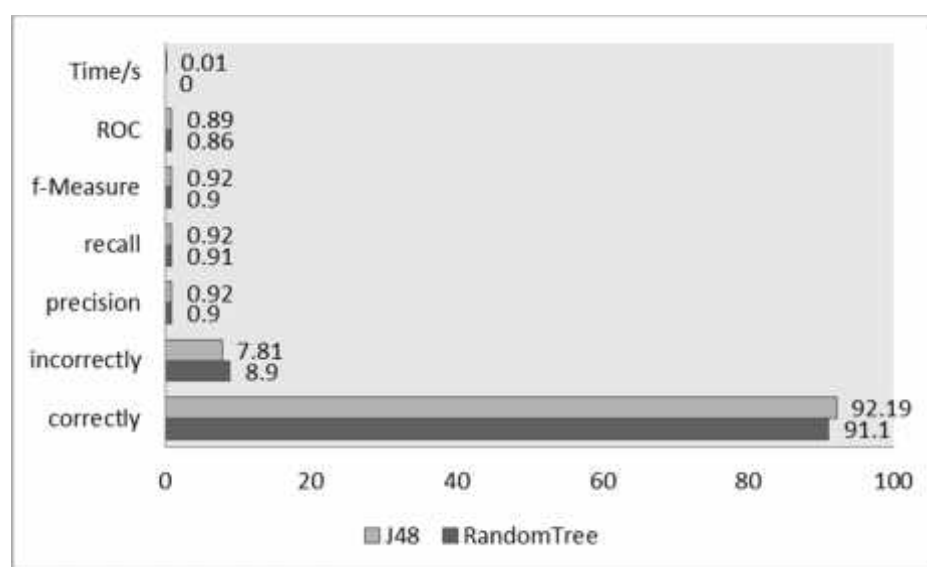
The two best accuracy of 10-fold cross validation for 1280 of instances were reached by the ID3 algorithm and Random Tree algorithm. Each algorithm gave the same precise accuracy of all parameters, accepting the time taken to build the model. ID3 needed 0.24 seconds longer to build the classification model compared to Random Tree algorithm. Otherwise, the accuracy that was performed by J48 algorithm for percentage split testing technique provided the best value for all classification techniques. It reached 92.19 % for accuracy and 0.01 second to build the model. The comparative of accuracy for two testing techniques for 1280 of instances can be seen in Figure 2.

Table 2. Accuracy of 10 fold Cross Validation Testing Techniques for 1280 of instances

No	Type of Tree	Time/s	Correctly	Incorrectly	Precision	Recall	F-Measure	ROC
1	AD Tree	0.16	90.00	9.70	0.89	0.90	0.89	0.86
2	BF Tree	2.00	91.00	8.90	0.90	0.91	0.90	0.86
3	Decision Stump	0.01	87.03	12.97	0.76	0.87	0.81	0.69
4	FT	1.13	91.25	8.75	0.90	0.91	0.90	0.84
5	Id3	0.24	91.10	8.90	0.90	0.91	0.90	0.86
6	J48	0.01	90.50	9.40	0.89	0.90	0.90	0.82
7	J48graft	0.08	90.50	9.40	0.89	0.90	0.90	0.82
8	LAD Tree	0.37	90.90	9.00	0.90	0.90	0.90	0.86
9	LMT	2.79	91.00	8.90	0.90	0.90	0.90	0.88
10	NB Tree	1.57	90.10	9.90	0.89	0.90	0.89	0.88
11	Random Forest	0.08	90.90	9.00	0.90	0.90	0.90	0.86
12	Random Tree	0.00	91.10	8.90	0.90	0.91	0.90	0.86
13	REP Tree	0.07	90.70	9.30	0.89	0.90	0.90	0.82
14	Simple Cart	1.23	90.70	9.30	0.89	0.90	0.90	0.82

Table 3. Accuracy of Percentage Split Testing Techniques for 1280 of instances

No	Type of Tree	Time/s	Correctly	Incorrectly	Precision	Recall	F-Measure	ROC
1	AD Tree	0.04	91.40	08.59	0.91	0.91	0.91	0.88
2	BF Tree	0.06	90.82	09.18	0.91	0.91	0.91	0.87
	Decision	0.00	87.69	12.30	0.77	0.88	0.82	0.76
3	Stump							
4	FT	0.44	92.19	07.81	0.92	0.92	0.92	0.90
5	Id3	0.03	90.82	09.18	0.91	0.91	0.91	0.87
6	J48	0.01	92.19	07.81	0.92	0.92	0.92	0.89
7	J48graft	0.02	92.19	07.81	0.92	0.92	0.92	0.89
8	LAD Tree	0.09	92.19	07.81	0.92	0.92	0.92	0.86
9	LMT	0.99	92.19	07.81	0.92	0.92	0.92	0.89
10	NB Tree	0.37	91.60	08.39	0.91	0.92	0.91	0.89
11	Random Forest	0.08	90.82	09.18	0.91	0.91	0.91	0.86
12	Random Tree	0.00	90.82	09.18	0.91	0.91	0.91	0.87
13	REP Tree	0.02	92.19	07.81	0.92	0.92	0.92	0.89
14	Simple Cart	0.34	90.82	09.18	0.91	0.91	0.91	0.88

**Figure 2.** Accuracy Comparative

6. Conclusion

This study have compared the performance of different tree classifier algorithm in WEKA toolkit using two kinds of testing techniques (10-fold cross validation & percentage split) for 1920 numbers of instances. From the study, we have shown that J48 algorithm with percentage split had the best percentage of accuracy based on correctly, incorrectly, precision, recall, f-measure, and ROC. Meanwhile, for the time taken to build the model, Random Tree algorithm had the best result (Table 4). J48 algorithm has built the prediction in Tree model to define the presence of learning motivation with 6 numbers of leaves and 11 for the size of the tree. The prediction showed that the presence of learning motivation on e-learning would be dependent upon a positive value from the attributes of r1 (the correctness of answering questions/assignment/exams/email), r2 (following the rules or standards), v3 (Speed of uploading assignments answer/ quizzes answer/ tests answer), and q (quantity /a length of the answer of assignment/exams/quizzes). The attribute of v2 which contained the speed of downloading e-learning material or assignment would not influence the presence of learning motivation.

Table 4. Classification Result based on Testing Technique.

Testing Method	Algorithm	Time/s	Correctly	Incorrectly	Precision	Recall	F-Measure	ROC
Tenfold cross validation	Random Tree	0.00	91.10	08.90	0.90	0.91	0.90	0.86
	Id3	0.24	91.10	08.90	0.90	0.91	0.90	0.86
Percentage Split	J48	0.01	92.19	07.81	0.92	0.92	0.92	0.89
	J48graft	0.02	92.19	07.81	0.92	0.92	0.92	0.89

7. Acknowledgments

Authors would like to say thank to:

1. Anonymous reviewers for their helpful comments
2. The Directorate of Higher Education (BPPDN Program) and STMIK “AMIKBANDUNG” for enabling the authors to obtain the opportunity and funding

8. References

- [1] Jan D. H, Dermot B, and Agnes M 2013 *What is Learning? On the Nature and Merits of a Functional Definition of Learning* Psychon Bull
- [2] Sandra G, and Bernard W 1996 *Theories and Principles of Motivation* University of California-Los Angeles
- [3] WEKA, University of Waikato, New Zealand
- [4] B. Rienties, D.Tempelar, P.V.D Bossche, W. Gijsselaers, and M. Segers, “The Role of Academic Motivation in Computer-Supported Collaborative Learning,” *Journal of Elsevier* 2009 in Human Behavior,
- [5] Hsiu-Feng Shih et al, 2013, “The relationship among tertiary level EFL students personality, online learning motivation and online learning satisfaction, *International Educational Technology conference, procedia – social and behavioral sciences* 103, pp 1152-1160
- [6] Dewi, A. Deshinta et al, 2011. A Computational Sistem Approach To Delevop Student’s Based Emotion Sistem: An Alternative Feedback Tool for Lecturers To Enhance Teaching and Learning, Faculty of Engineering and Information Technology, INTI International University, Nilai, Malaysia
- [7] A.Darouich, F. Khoukhi, and K. Douzi, “Mining Fuzzy Motivation Indicator in Learning Environment through Human Computer Interaction,” *Science and Information Conference*, London-UK, 2013 Maastricht University, Limburg-Netherland.
- [8] Christina J, Arry A.A , Husni S.S , Iping S. *Measurement of Learning motivation in Electronic Learning*, International Conference on Information Technology Systems and Innovation (ICITSI), Bandung-Bali, Indonesia. 2015.
- [9] Koggalahewa D.N, Amararachi, Pilapitiya, Geegange. *Semantic Self Learning and Teaching Agent (SESLATA)*. The 8th International Conference on Computer Science & Education (ICCSE 2013). Colombo-Sri Lanka. 2013
- [10] Taruna S, Mrinal P.An Empirical Analysis of Classification Techniques for Predicting Academic Performance, *IEEE International Advance Computer Conference-IACC*.2014