

# Mining Educational Data to Analyze the Student Motivation Behavior

Kunyanuth Kularbphettong, Cholticha Tongsiri

**Abstract**—The purpose of this research aims to discover the knowledge for analysis student motivation behavior on e-Learning based on Data Mining Techniques, in case of the Information Technology for Communication and Learning Course at Suan Sunandha Rajabhat University. The data mining techniques was applied in this research including association rules, classification techniques. The results showed that using data mining technique can indicate the important variables that influence the student motivation behavior on e-Learning.

**Keywords**—association rule mining, classification techniques, e-Learning, Moodle log Motivation Behavior

## I. INTRODUCTION

WITH the rapidly growth of the Internet, the web-based educational system has been increasingly used as an important tool to support learners and teachers. There are many benefits for information sharing and collaboration between learners and teachers in a course. Learners can take a web-based class to enhance their knowledge at any time and any place and teachers can easily create their online classes and monitor student's performance as well. Moodle is a popular Learning Management System (LMS) that supports educators to create the effective online courses. However, it did not provide the function to access and estimate learner's motivation behavior. Also Student's logs in a Moodle can show students' interactions like reading, writing, taking exam, and doing various tasks [1]. With a huge of accumulated information daily from e-Learning course, it is very difficult to analyze this data manually. Although there are some tools that help to report useful information which in turn is very valuable for analyzing student's pattern behavior [2], they do not offer specific features teacher need to track and evaluate all the students' activities in class [3].

Data Mining Techniques is the promising methodology to extract valuable information in this objective. Data Mining can analyze relevant information results and produce different perspectives to understand more about the students' activities so as to customize the course for student learning. For this paper, we used the data set of the Information Technology for Communication and Learning Course at Suan Sunandha Rajabhat University.

Kunyanuth Kularbphettong is with Computer Science Program, Suan Sunandha Rajabhat University, Bangkok, Thailand (phone: 662-150-1169; e-mail: kobkulriss@gmail.com).

Cholticha Tongsiri is Faculty of Information and Communication Technology, Silpakorn University, Bangkok, Thailand (phone: 662-233-4995; e-mail: c\_tongsiri@hotmail.com).

The data is composed of demographic information of each student, information in this class and grade student obtained. After data preparation process, data mining techniques are applied to discover the valuable knowledge for analysis student motivation behavior. And the rest of this paper is organized as follows. Section 2 reviews about the related literatures and the related methodologies used in this work. Section 3 presents the implementation based on the purposed data mining techniques. In section 4 the result and discussion is presented. Finally, we conclude the paper with future research issues in section 5.

## II. RELATES WORKS AND THE METHODOLOGIES

In this section, we illustrate the literature search and the specified methodologies used in this project.

### A. Relates Works

A literature search shows that most of the related researches have deployed data mining techniques to analyze student's learning behaviors by following this: According to C. Romero et al [4], the research was shown the usefulness of the data mining techniques in course management system and the rules can help to classify students and to detect the sources of any incongruous values received from student activities. K. Waiyamai applied data mining approach to develop new curricula and to assist engineering students for choosing an appropriate major [5] and B. Minaei-Bidgoli et al [6] presented an approach to classify students in order to predict their final grade based on features extracted from logged data in an education web-based system. Data mining techniques like association rule mining were applied in [7],[8] to extract the patterns and to evaluate the activities of on line course and classification and association rule mining algorithms are discussed and demonstrated in [9]. Also there are many researches that have been investigated in the on-line learning environment. For example, West et al investigated impact of learning style on e-learning by using Statistics [10] and Kerdprasop et al used Rule induction rough set to Classify student knowledge level [11].

### B. The Methodologies

Data Mining is the data analyzing approach from different perspectives to summarize the useful information results. The data mining process uses many principles as machine learning, statistics and visualization techniques to discover and present knowledge in an easily comprehensible form. There is another definition as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [12], [13].



TABLE I  
IMPORTANT ATTRIBUTES

Name	Description
PreTest_number	Identification number of pretest
PostTest_number	Identification number of posttest
Result-PreTest_number	Mark obtained from pretest
Result-PostTest_number	Mark obtained from posttest
Time-of-PreTest_number	Time spent on pretest
Time-of-PostTest_number	Time spent on posttest
SumofPreTest	Total mark obtained from all pretest
SumofPostTest	Total mark obtained from all posttest
Time-of-viewing –material	Total time spent on viewing material
Time-of-upload –material	Total time spent on upload material
number-of-Attendance	Number of Attendance in class
Assignment_Number	Identification number of assignment
Result_Assignment_Number	Mark obtained from assignment
InClassTest_number	Identification number of in class test
Result_InClassTest_number	Mark obtained from in class test
Project_score	Mark obtained from project
Midterm-score	Mark obtained from Midterm
Final-score	Mark obtained from Final
Grade	Final mark obtain from this class

## IV. RESULTS OF EXPERIMENTAL

## A. Apriori Algorithm

Apriori algorithm has been used for this research and evaluated with a minimum support of 0.2 and a minimum confidence of 0.9 to presents significant relationships between the activities of students on this class and their final scores. Fig 3 shows examples from the results of Apriori algorithm.

Best rules found:

1. Midium=Very Low 409 ==> Count-Test-P=Very Low 397 conf:(0.98)
2. Count-P-test=Height 469 ==> Count-Test-P=Very Low 448 conf:(0.98)
3. Count-P-test=Height Midium=Low 341 ==> Count-Test-P=Very Low 324 c
4. Count-P-test=Very Height Count-Test-P=Very Low Grade=High 347 ==> Cou
5. Count-P-test=Very Height Grade=B+ 406 ==> Count-assign=Very Height 38
6. Count-Test-P=Very Low Grade=B+ 374 ==> Count-assign=Very Height 353
7. Grade=High 433 ==> Count-assign=Very Height 408 conf:(0.97)
8. Count-assign=Very Height Grade=High 408 ==> Count-P-test=Very Height
9. Grade=High 433 ==> Count-P-test=Very Height 406 conf:(0.97)
10. Final=Medium Grade=High 341 ==> Count-assign=Very Height 319 conf:
11. Count-assign=Height 441 ==> Count-Test-P=Very Low 412 conf:(0.97)
12. Count-assign=Very Height Count-Test-P=Very Low Grade=High 353 ==> Cou
13. Count-assign=Very Height In-Class=Very Height Project=Very Height 371
14. Final=Medium Grade=High 341 ==> Count-P-test=Very Height 317 conf:

Fig. 3 Results of Apriori algorithm

From the results of Apriori approach, there are many unnecessary rules teachers are not interested in. Therefore, for this research, the selected rules are explained in the Fig.4, based on decision of teacher.

Discovered rules	Conf.
Count-P-test=Very High Final=Medium Count-assign=Very High ==> Grade= Very High	0.98
Final=Medium Count-assign=Very High ==> Grade= Very High	0.98
Count-assign=Very High In-Class=Very High Count-P-test=Very High ==> Grade= Very High	0.98
Count-P-test=High Project=High Count-Test-P=Very Low ==> Grade= Very High	0.98
Count-assign=Very High Final=Medium Count-P-test=Very High ==> Grade= Very High	0.98
In-Class=Very High Count-P-test=Very High ==> Grade= Very High	0.97
Count-P-test=Very High In-Class=Very High Count-assign=Very High ==> Grade= Very High	0.97
Count-P-test=Very High Count-assign=Very High ==> Grade= Very High	0.97
In-Class=Very High Count-assign=Very High ==> Grade= Very High	0.97
Count-Test-P=Very Low Medium=Low Count-assign=Very High ==> Grade= High	0.97
Count-P-test=Very High Medium=Low Count-assign=Very High ==> Grade= High	0.97
Mid=Low Count-assign=Very High ==> Grade= High	0.96
Mid=Very Low Count-Test-P=Very Low ==> Grade= High	0.96
Final=Medium Count-assign=Very High Count-P-test=Very High ==> Grade= Very High	0.96
Count-P-test=Very High In-Class=Very High Count-assign=Very High ==> Grade= High	0.95

Fig. 4 Results of selected rules

“Grade”, is significantly attributed to student motivation behavior. Accordingly, it was being discretized in to five levels: VERY LOW, LOW, MEDIUM, HIGH and VERY HIGH. Based on the derived testing result, 30% of the data set used for training, the result is depicted in Figure 5.

Student Motivation Behavior Levels	Testing set	Accuracy	
		Number of Record	Percentage
Very High	224	187	83.5
High	284	254	89.44
Medium	164	125	76.22
Low	112	87	77.68
Very Low	117	78	66.67
Total	901	731	81.13

Fig. 5 Results of testing mined rules

## B. Decision Tree Algorithms

ID3 and J48 are utilized in this research and Fig 6(a) and (b) present the results of these algorithms.

Id3

```

Count-assign = Very Height
| In-Class = Very Height
| | Midium = Low
| | | Count-P-test = Very Height
| | | | Final = Low
| | | | | Count-Test-P = Very Low: High
| | | | | Count-Test-P = Very Height: High
| | | | | Count-Test-P = Height: High
| | | | | Count-Test-P = Medium: Medium
| | | | | Count-Test-P = Low: High
| | | | Final = Very Low: null
| | | | Final = Medium
| | | | | Count-Test-P = Very Low
| | | | | Project = Very Height: Very High
| | | | | Project = Height: High
| | | | | Project = Very Low: null
| | | | | Project = Medium: High
| | | | | Project = Low: null
| | | | Count-Test-P = Very Height: Very High
| | | | Count-Test-P = Height: null
| | | | Count-Test-P = Medium: High
| | | | Count-Test-P = Low: null
| | | | Final = Height
| | | | | Project = Very Height: Very High
| | | | | Project = Height: Very High
| | | | | Project = Very Low: null
| | | | | Project = Medium: null
| | | | | Project = Low: null
| | | | Count-P-test = Medium: Medium
| | | | Count-P-test = Very Low: null
| | | | Count-P-test = Height

```

Fig. 6 (a) Results of ID3

```

J48 pruned tree
-----
Count-assign = Very Height
Count-P-test = Very Height
|
|   In-Class = Very Height
|   |
|   |   Final = Low: High (5.0/1.0)
|   |   |
|   |   |   Final = Very Low: Very High (0.0)
|   |   |   |
|   |   |   |   Final = Medium: Very High (31.0/14.0)
|   |   |   |   |
|   |   |   |   |   Final = Height: Very High (5.0/2.0)
|   |   |   |   |   |
|   |   |   |   |   |   In-Class = Height
|   |   |   |   |   |   |
|   |   |   |   |   |   |   Midium = Low
|   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   Final = Low
|   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   Project = Very Height: Medium (2.0)
|   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   Project = Height: Medium (6.0/2.0)
|   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   Project = Very Low: Medium (0.0)
|   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   Project = Medium: Medium (1.0)
|   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   Project = Low: Medium (0.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   Final = Very Low: High (0.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Final = Medium: High (11.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Midium = Medium: Very High (7.0/2.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Midium = Height: Medium (0.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Midium = Very Low: Low (3.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   In-Class = Medium
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Final = Low: Low (2.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Final = Very Low: High (0.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Final = Medium
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Midium = Low: Medium (3.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Midium = Medium: Medium (2.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Midium = Height: Medium (0.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Midium = Very Low: Medium (0.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   Final = Height: Medium (0.0)

```

Fig. 6 (b) Results of J48

The table II shows that ID3 has the correctly classified accuracy of 83.9161% and J48 has the accuracy of 68.5315%. And the table III presents the classifier accuracy by class.

TABLE II  
CLASSIFIER ACCURACY

Algorithm	Correctly Classified Instance	Incorrectly Classified Instances
ID3	83.9161 %	16.0839 %
J48	68.5315 %	31.4685 %

TABLE II  
CLASSIFIER ACCURACY BY CLASS

Algorithm	Class of Student Motivation Behavior	TP Rate	FP Rate
ID3	VERY HIGH	0.571	0.007
	HIGH	0.64	0.059
	MEDIUM	0.8	0.073
	LOW	0.96	0.034
	VERY LOW	0.75	0.008
J48	VERY HIGH	0.429	0.015
	HIGH	0.88	0.136
	MEDIUM	0.4	0.016
	LOW	0.76	0.051
	VERY LOW	0.65	0.057

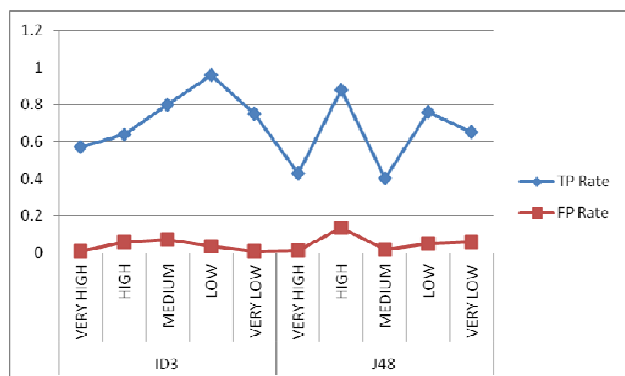


Fig. 7 Results of classifier accuracy by class

## V.CONCLUSION AND FUTURE WORK

In this paper, we show the data mining techniques can be useful to model the student behavior outcomes and the preliminary result presents a promising progress in this prototypes model for the ongoing improvement of e-Learning course.

This model can be beneficial to similar courses to share and discover students' motivation behavior. However, in term of the future experiments, we are looking forward to research about other data mining techniques to enhance this project and also apply the tool to help teachers in their class.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the financial subsidy provided by Suan Sunandha Rajabhat University.

## REFERENCES

- [1] J. Mostow, J. Beck, H. Cen, A. Cuneo, E. Gouvea, and C. Heiner, "An educational data mining tool to browse tutor-student interactions: Time will tell", In Proceedings of the Workshop on Educational Data Mining, Pittsburgh, USA (pp. 15–22), 2005.
- [2] E. García, C. Romero, S. Ventura, and C. Castro, "Using Rules Discovery for the Continuous Improvement of e-Learning Courses," *Lecture Notes in Computer Science*, 2006, Volume 4224/2006, 887-895.
- [3] M. E. Zorrilla, E. Menasalvas, D. Marin, E. Mora, and J.Segovia, "Web usage mining project for improving web-based learning sites", In Web Mining Workshop. Cataluna pp. 1-22, 2005.
- [4] C. Romero, S. Ventura, E. García, "Data mining in course management systems: Moodle case study and tutorial" *Computers & Education*, Volume 51, Issue 1, August 2008, pp. 368–384.
- [5] Waiyamai, K. "Improving Quality Graduate Student by Data Mining". Department of Computer engineering. Faculty of Engineering. Kasetsart University, Bangkok Thailand. 2003.
- [6] B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer and, W. F. Punch. "Predicting student performance: an application of data mining methods with the educational web-based system LON-CAPA" In Proceedings of ASEE/IEEE Frontiers in Education Conference, Boulder, CO: IEEE, 2003.
- [7] H.H. Hsu, C.H. Chen, and W.P. Tai, "Towards Error-Free and Personalized Web-Based Courses", In: The 17th International Conference on Advanced Information Networking and Applications, AINA'03. March 27–29, Xian, China, pp. 99–104, 2003.
- [8] A. Kumar, "Rule-Based Adaptive Problem Generation in Programming Tutors and its Evaluation", In: The 12<sup>th</sup> International Conference on Artificial Intelligence in Education. July 18–22, Amsterdam, pp. 36–44, 2006.
- [9] F. Berzal, J.C. Cubero, N. M. Sánchez, J.M. Serrano, and A. Vila, "Association rule evaluation for classification purposes" *Actas del III Taller Nacional de Minería de Datos y Aprendizaje*, TAMIDA2005, pp.135-144 ISBN: 84-9732-449-8, 2005.
- [10] W. West, B.R.S. Rosser, S. Monani, and L. Gurak, "How Learning Styles Impact ELearning: a Case Comparative Study Of Undergraduate Students Who Excelled, Passed Or Failed An Online Course", In Scientific/Technical Writing. E-learning, pp. 534–543, 2006.
- [11] N. Kerdprasop, N. Muenrat, and K. Kerdprasop, "Decision Rule Induction in a Learning Content Management System" *Proceedings of World Academy of Science, Engineering and Technology*, pp.77–81, 2008.
- [12] U. M. Fayyad, G. Pitatesky-Shapiro, P. Smyth, and R. Uthurasamy, "Advances in Knowledge Discovery and Data Mining", AAAI/MIT Press, 1996.
- [13] W. Frawley, G. Piatetsky-Shapiro, and C. Matheus, "Knowledge Discovery in Databases: An Overview". *AI Magazine*, Fall 1992, pp. 213-228.
- [14] R. Agrawal, T. Imielinski, and A.N. Swami, A. N., "Mining association rules between sets of items in large databases", In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216, 1993.
- [15] Z. Qiankun, "Association Rule Mining: A Survey, Technical Report", CAIS, Nanyang Technological University, Singapore, 2003
- [16] R. Agrawal, and R. Srikant, "Fast algorithms for mining association rules", In Proc. 20<sup>th</sup> Int. Conf. Very Large Data Bases, VLDB, pp. 487-499, 1994.
- [17] H. Edelstein, "Introduction to Data Mining and Knowledge Discovery", Third Edition. Two Crows Corporation, Potomac, MD, USA, 1999.

- [18] Parr Rud, O. "*Data Mining Cookbook. Modeling Data for Marketing, Risk, and Customer Relationship Management*". John Wiley & Sons, Inc.; 2001.
- [19] S. Mitra, and T. Acharya., "*Data Mining. Multimedia, Soft Computing, and Bioinformatics*". John Wiley & Sons, Inc. Hoboken, New Jersey; 2003.
- [20] J. R. Quinlan, "*Introduction of decision tree*", Journal of Machine learning", pp. 81-106, 1986.
- [21] L. Hyafil, and RL. Rivest, "*Constructing Optimal Binary Decision Trees is NP-complete*" , Information Processing Letters, Vol. 5, No. 1. (1976), pp. 15-1 R. Kohavi and J. R. Quinlan. Decision-tree discovery.
- [22] A. Kumar Sharma and S. Suruchi, "*A Comparative Study of Classification Algorithms for Spam Email Data Analysis*", International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 5, pp. 1891-1895. May 2011
- [23] <http://www.cs.waikato.ac.nz/ml/weka/>
- [24] E. García, C. Romero, S. Ventura, C. Castro, and T. Calders, "*Chapter 7: Association Rule Mining in Learning Management Systems.*" In: Hadebook of Educational Data Mining, Taylor&Francis Group, 2010.

**Kunyanuth Kularbphettong** received the B.S. degree in Computer Business, M.S. degree in Computer Science, and Ph.D degree in Information Technology. Her current research interests are in Multi-agent System, Web Services, Semantic Web Services, Ontology and Data mining techniques.

**Cholticha Tongsiri** is interested in Web-Based Educational Learning, and Data mining techniques. She is studying MS. In Management Information System.