MINING YUC[1] STUDENTS LEARNING BEHAVIOR IN MOODLE SYSTEM USING DATAMINING TECHNIQUES

Musa I. Swedan

Department of Management Science, [1]Yanbu University College, KSA.

Abstract— **Recently, educational dataMining has become an interesting research area to discover and extract hidden knowledge of students' patterns from educational environment. In this work we have two purposes. First to apply the dataMining techniques, namely, Classification, Clustering and Association Rule mining algorithms on the data stored on the e-learning (Moodle system) database to extract knowledge's that help to understand students' behaviour's and patterns, improve teaching practice, and to assist educators to evaluate and improve e-learning system [1]. The secondly purpose is to develop a statistical tool (a Web application) for the educators in YU College to monitor his/her students learning behavior. It will enable the instructors for instance, to monitor the number of assignments taken, the number of quizzes taken, the number of forum post and/or read by students, etc. This kind of knowledge's can helps the instructors to make an immediate reaction to the way the students interact with the courses activities in the e-learning, and to create an effective educational environment.**

In this paper, RapidMiner (v5.0) dataMining tool were used for mining the data from the Moodle system for a various courses taken by Management Science, Computer Science and Applied Linguistic students at YUC during 2011/2012 semester.

Keywords— **Educational DataMining (EDM), DataMining Algorithms, Moodle System, Student Behavior, RapidMiner, SMoodle System.**

I. Introduction

A. *Data Mining*

Data Mining (DM) or knowledge discovery in database (KDD) is the automatic extraction of hidden and interesting patterns from large data collection [2]. DM comes as a tool to assist humans in extracting useful information from the rapidly growing volumes of digital data [3]. DM use advanced techniques and methods to discover knowledge, some the most useful DM methods are: statistics, visualization, clustering, classification, associations rule mining, text mining, etc.

In the educational system both in *Traditional classroom* and *Distance education*, data Mining methods are used to help teachers to improve the learning environment in the educational institution, which eventually will effect positively on the learning experience of students.

In traditional education teachers attempts to *enhance instructions* by monitoring student's learning processes and analyzing their performance by paper records and observation, students attendance, course information, etc. however, when students work in electronic environments (i.e. in distance learning) the students are separated by time and space from teachers. So the methods of monitoring the student information in this environment are different. Web-based learning environment are able to record most if not all the learning behaviors of the students and educators, and are hence able to provide a huge amount of learning data [1].

In this paper the e-learning system also known as Learning Management System (LMS) is used here. E-learning contains large amount of data which could be extracted to generate information for analyzing students and educator's behavior. E-learning system offers a virtual communication between the educators and students, sharing resources, producing course content (in any type format: documents, multimedia), conducting online tests, synchronous learning (such as forums, chats and news). *Moodle* system [8] is an example of an open source e-learning system and it will be the bases for this work.

B. *Educational Data mining (EDM)*

EDM was defined by Baker [5] "as the area of scientific inquiry centered around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to better understand students and the settings which they learn in". In another words EDM concerns with developing methods that discover knowledge from data come from educational environment. The data can be collected from historical and operational data reside in the database of educational institutes. And the data reside on the e-learning systems database. EDM have the following categories [1]: statistics and visualization, web mining (clustering, classification, association role mining and sequential pattern mining), and text mining.

This paper is arranged in the following way: Section II describes the data preparation and preprocessing steps. Section III shows the experimental results for dataMining algorithms. In section IV, SMoodle is introduce, its structure and components.

II. Data preparation and pre-processing

In general e-learning dataMining process consists of four steps as follows:

1) *Collect data*. Targeted dataset (i.e. activities perform the uses) are stored in the Moodle database.
2) *Preprocess the data*. Transform the data into appropriate format. Help to improve the accuracy, efficiency and scalability of DM algorithms. This includes data cleaning and processing, data reduction

and transformation. To increase the interpretation and comprehensibility discretized is used to transform the numerical attributes to categorical ones. For instance number of assignments students submitted for particular course in e-learning can be categorize into Zero, Low, Medium or High.

3) *Apply data mining*. In this paper clustering, classification and association rules algorithms are applied to discover the knowledge's and patterns of interest. Different dataMining tools are used to apply dataMining algorithm. RapidMiner, Weka, and keel are example of open source DM tools. RapidMiner is used in this paper.
4) *Interpreting mined patterns*. The results obtain from dataMining tools can be used be the educators to make decisions about his/her students in a course to improve learning environment.

DataMining in the e-learning system is an iterative cycle [1] in which the mined knowledge should enter the loop to enhance leaning as a whole, not just turning row data into knowledge, but also make the decision of the revised knowledge. Fig. 1 shows the four steps for preprocessing Moodle database.

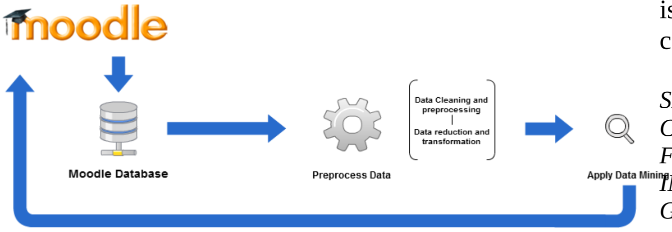

moodle
Moodle Database
Preprocess Data
Data Cleaning and preprocessing
Data reduction and transformation
Apply Data Mining

Fig. 1 Moodle database preprocessing process

III. EXTERMINATION AND RESULTS

A. *MOODLE DATABASE*

Moodle data are stored in single relational database manly consist of teachers and students activities, courses data, etc. latest moodle release (Moodle 2.3.2+) has about 302 interrelated tables. Some of the most important Moodle tables are listed in Table1.

TABLE I
IMPORTANT MOODLE TABLES

Table Name	Description
yuc_log	Contains every users activities
yuc_user	Contains information about all users
yuc_assignment	Assignment information
yuc_assignment_submissions	Assignments submitted by students
yuc_quiz	Information about all quizzes
yuc_quiz_grades	Information about quiz grades
yuc_quiz_attempts	Store quiz attempts for various students
yuc_forum	Information about all forums
yuc_forum_discussions	All Forums' discussions
yuc_forum_posts	All posts to the forum

The stored data in the Moodle database need to transforms into a formats that can recognize by the data mining tools and algorithms. To accomplish this task, the data preprocessing have to apply on the database (see section II), and later on to create the *summarization table*. It contains all the information of the involved students' activities. And it will be used by the DM algorithms to extract required knowledge's.

B. *SUMMARIZATION TABLE*

Summarization table created by transform the original tables in the Moodle into a format a particular dataMining tools and algorithms can work with it. The transformation process consists of the following steps:

1) Selecting the data

Yanbu university e-learning system has approximately 3507 students in 176 courses. Of course not all the courses are eligible. Only courses with higher activities (i.e. the course with at least assignments, quizzes and forums) are chosen. Eventually, only 22 courses and 1883 students are involved in this study. Simple query statements written in MySQL (see query example below) used to check wither the courses have an assignments or not. The same query can apply for quizzes and the forums with slight changes. Courses with zero query result, that is, the courses with no assignment, forum or quiz are not considered and will be omitted from the experimentation.

```
SELECT c.fullname /*course name*/, a.course /*course id*/,
COUNT(*) AS "assignment count"
FROM yuc_assignment a
INNER JOIN yuc_course c ON c.id = a.course
GROUP BY a.course
```

2) Information integration

Required data in the Moodle database are spread over several tables, so the relevant data are needed to integrate in one place. Table 2 shows summary table, each attributes in this table has a summary about all the activities perform by each student in the course.

TABLE 2
SUMMARIZATION TABLE (NUMERICAL VERSION)

Attribute Name	Description
course *(cn)*	course number/code
assignment_number *(ac)*	Number of assignments taken by the student
assignment_total_time *(at)*	Total time spent on assignment
quiz_number *(qc)*	Number of quizzes taken
number_quiz_passed *(qp)*	Number of quizzes passed
number_quiz_failed *(qf)*	Number of quizzes failed
quiz_total_time *(qt)*	Total time spent on quizzes
forum_post *(fp)*	Number of forum posts
forum_read *(fr)*	number of forum reads
resource_view *(mv)*	Number of course materials views
final_grade *(fg)*	Student final grade

To create the summary table, several SQL queries statements to the Moodle database are performed to obtain information for students' activities. For example, in order to query the total number of assignments taken by the students on particular course. The following SQL statement can accomplish this task:

```
SELECT COUNT(sub.userid) AS "number of assignemnts"
FROM yuc_assignment AS a, yuc_assignment_submissions AS
sub, yuc_user AS u
WHERE a.course = 9 /*Course ID on the Moodle*/
AND a.id = sub.assignment
AND u.id = sub.userid
GROUP BY (userid)
```

Table 3 shows a snapshot of the summary table with a numeric values generated from the Moodle database. The first column is the *course name* (cn) and the last column is the *final grade* (fg) of the student. From the row no 2, the student submitted only 3 assignments, the total time needed for submission is 22 hours, the number of quizzes taken are 3. And his/her final grade is 80%. For the rest of columns description, see table 2.

TABLE 3
SUMMARY TABLE WITH NUMERICAL VALUES

cn	ac	at	qc	..	fg
CS001	0	0	1	..	0
CS001	3	22	3	..	80
CS001	0	0	3	..	60
CS001	0	0	1	..	0
CS001	0	0	3	..	82
…	…	…	…	..	..

3) *Data discretization*

The purpose of discretization step is to transform numerical data into categorical classes (or finite set of intervals). The total accuracy for the dataMining algorithms will increase when discretized data are used [11]. All the numerical values of the summarization table (see table 3) categorized into 4 labels (see table 4) using equal-width method (this method divides the range of the attribute into a fixed number of intervals of equal length) [6] the 4 labels are: *Zero* (for instance, the student's did not make any assignments submission), *Low*, *Medium* and *High*, except the final grade attribute has 3 labels: *Fail*, *Pass*, and *Excellent* label, created using *manual method*, that is the labels assigned for the students, if the final grade <60%, final grade >=60%, and final grade >=90%, respectively. No changes on the course attribute.

TABLE 4
SUMMARIZATION TABLE (CATEGORICAL VERSION)

Attribute Name	Values
Course	CS001, MIS342, MIS102, ENG202
assignment_number	Zero, Low, Medium, High
assignment_total_time	Zero, Low, Medium, High
…	…
final_grade	Fail, Pass, Excellent

TABLE 5
SUMMARY TABLE WITH CATEGORICAL VALUES

cn	ac	at	qc	..	fg
CS001	Zero	Zero	Low	..	Fail
CS001	Medium	High	Medium	..	Pass
CS001	Zero	Zero	Medium	..	Pass
CS001	Zero	Zero	Low	..	Fail
CS001	Zero	Zero	Medium	..	Pass
…	…	…	…	..	…

C. *APPLYING DATA MINING ALGORITHMS AND INTERPRETING RESULTS*

One of the most well-known dataMining tools on the data mining community is the RapidMiner [7]. It is contains almost of 600 operators, enormous options of data transformation, dozens of modeling operations, several of data input operators (CSV file, Excel Sheet, Database table, ARFF, SPSS), etc. which makes it the most comprehensive solution for dataMining operations. RapidMiner or Rapid-I will be used to apply the algorithms that we choose for mining students patterns.

Before starting applying dataMining algorithms it is worth to find out the weights for each attributes in table 2. Information gain (IG) method [10] used for attributes weighting, IG gives a good indication for how much the students are involved in a particular activity, see table 6.

TABLE 6
WEIGHTING WITH INFORMATION GAIN

Attribute	Information Gain
assignment_number	0.38
assignment_total_time	0.23
quiz_number	0.33
number_quiz_passed	0.79
number_quiz_failed	0.54
quiz_total_time	0.22
forum_post	0.00
forum_read	0.03
resource_view	1.00

In the table 6 the *resource view* attribute has the highest weight (with IG=1.00), this good indication that the moodle students are viewing the courses materials frequently. *Number of quiz passed* attribute come in the second place (with IG=0.79). *Forum post* and *forum view* have the lowest weights, with IG=0.00 and IG=0.03, respectively, this indicate that the students have few involvement in the courses forums, the educators can use these numbers to find out the strength and weakness of his/her students.

1) *Classification algorithm*

Classification is a supervised classification used to predict class label. For instance we can classify students into different groups with equal final grade. With this technique we can predict failing students before the end of the semester, thus the educators can make an immediate reaction to adjust students' performance before the final exams.

The C4.5 [10] algorithm is used to characterize the students who passed or failed the course. With C4.5 we can define a set of logical rules (IF-THEN-ELSE rules) from the decision tree that can show interesting information about the classification of the students. See Fig. 2 for Decision tree.

In the decision tree, the student directly classified as Fail if the students have either low or zero number of quizzes taken, and the students with high number of quizzes taken directly classified as Pass.

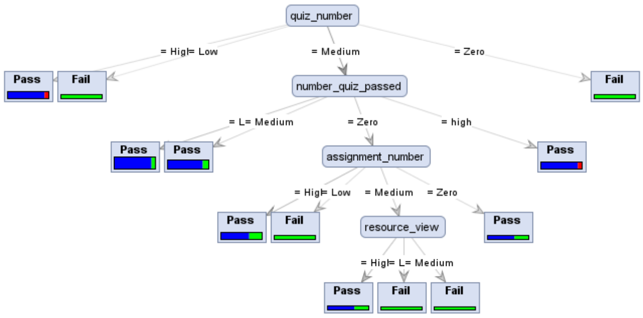


Fig. 2 Decision tree

It is important that the mining models be validated by understanding their quality (i.e. the accuracy) and characteristic before they are deployed into a production environment. There are several ways to measure this performance by comparing *predicted label (testing data)* and *true label (or training data)*. The usual way to estimate performance is therefore, to split the labeled dataset into a training set and a test set, which can be used for performance estimation. RapidMiner supports several operations for evaluating the accuracy, namely, Split Validation, X-validation, Bootstrapping validation, etc. fig. 3 shows the X-validation operation inside it the training and testing area (on the top). The decision tree is used for training the data. In the testing area the performance such as: accuracy, precision, recall, etc. are calculated. Confusion Matrix (CM) is also generated [12]. The CM displays the number of correct and incorrect predictions made by the model compared with the actual classifications in the test data.

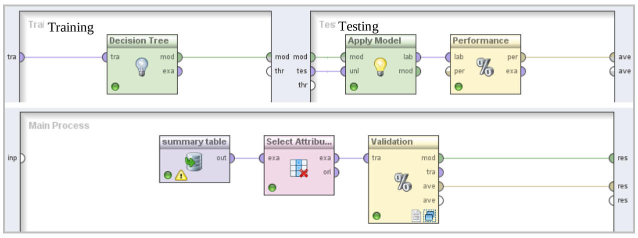


Fig. 3 X-Validation operation

After applying the x-validation operation the total accuracy have **78.22%.** In the Confusion Matrix the Pass label has recalled equal to 87.88% (i.e. 87.88% of the testing data are correctly labeled or classify). Fail label has recalled equal to **56.52%** and **0.00%** recall for excellent label.

2) Clustering algorithm

Clustering is a process of grouping objects into class of similar objects. In e-learning it has been used for finding groups of students with similar learning characteristics. For instance we can distinguish *active* students from *non-active* students according to their activities in the e-learning. *K-means* algorithm is going to use in this paper. It is one of the most popular and simplest clustering algorithms.

To apply k-means algorithm summarization table with numeric values (see table 3) are used. Of course labeled attribute (i.e. the final grades attribute) should omit. In table 7 we have 3 clusters of students. The educator can use this information to group students into different categories, for example, very active, active, low active students. In the table the students with highest assignments taken are in cluster 1 (with mean/mode ≈ 9.47), the lowest numbers of assignments found are in cluster 0 (with mean/mode ≈ 0.63). The highest numbers of students post and read the forums are in cluster 0 (with mean/mode ≈ 0.008, 0.46) and the lowest numbers of forums post and read are in cluster 1 (with mean/mode ≈ 0.0, 0.176).

TABLE 7
IMPORTANT MOODLE TABLES

Attribute	Cluster 0	Cluster 1	Cluster 2
assignment_number	0.63	9.47	4.053
assignment_total_time	52.77	5624.16	1003.09
quiz_number	1.19	0.88	2.61
number_quiz_passed	0.71	0.70	1.76
number_quiz_failed	0.48	0.176	0.84
quiz_total_time	19.54	244.06	14.56
forum_post	0.008	0.0	0.0
forum_read	0.46	0.176	0.36
resource_view	55.96	213.29	149.70

3) Association rule mining algorithm

Association rules searches for interesting relationships among items in a given dataset [13]. An association rule is an implication X→Y, where X and Y are disjoint item sets. The intuitive meaning of such a rule is that when X appears, Y also tends to appear. Each rule is accompanied by two measures, confidence and support. Apriori, FP-growth, etc. are example of association rule mining algorithms.

Applying association rule algorithm on the summary table with categorical values used for discovering the interesting relationships from student's usage information in order to provide feedback of course educator, for finding out the relationships between each pattern of learners behaviors, and determine what most interests the e-learning users.

IV. STATISTIC MOODLE (SMOODLE)

When this project started one purpose was in mind, is to develop an application that makes the Moodle user (specially the educator's) experience a new ways to comprehend his/her students learning behaviors in elegant and user-friendly ways.

SMoodle or Statistic Moodle is a web application written in PHP and JavaScript languages. It is works as a front-end for the Moodle System.

All activities performed by the students and teachers in the Moodle system are stored in a MySQL database (see fig. 4). SMoodle retrieve these data using PHP engine, and presented it to SMoodle users in more meaningful ways to support educators teaching process.

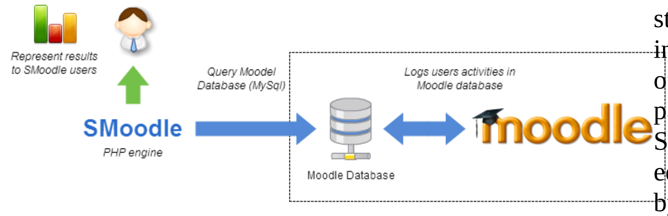

Represent results to SMoodle users
SMoodle
PHP engine
Query Moodel Database (MySql)
Logs users activities in Moodle database
Moodle Database
moodle

Fig. 4 SMoodle structure

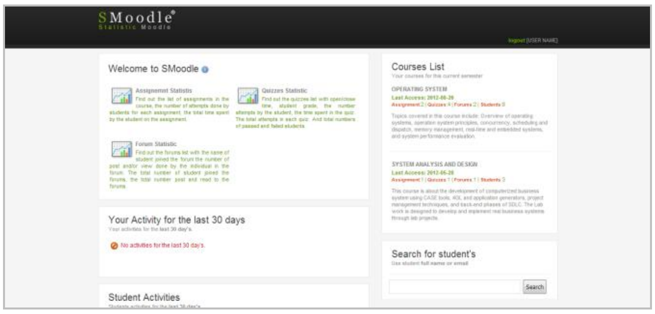
Moodle System has a various type of reports to show the students activities in a course (e.g. submitted assignments report, quizzes grades report, etc.), but these reports are hard to find and presented in a way make the interpretation of information from it hard for the normal educators. Eventually, it will effect on the students' outcome and the decisions taken by the educators as a whole.

SMoodle use the bar chart to visualize the retrieve data. To create the visual graph, *Ext-JS* [9] framework is used. Ext-JS is a JavaScript application framework that works in every browser. It enables developers to create the best cross-platform applications using nothing but a browser. Ext-JS has large collection of useful APIs, but only the chart API is used to serve our purpose.

A. *SMOODLE TEACHER LEVEL*

Teacher level consists of four main blocks: (1) *Assignment block*, (2) *Quiz block*, (3) *Forum block* and (4) *student block*. Each one of these blocks consists of a set of functions to support the educators of the Moodle. Assignment block provide statistical functions, such as, the list of assignments available in the course, the number of successful attempts performed by students in assignment, the total time spent on assignment by the students, etc.

In the Quiz block the educators can find out the available quizzes list along with quiz started and close date/time, quiz final grades, the number of attempts per student, and the time student spent in quiz. The total attempts in the quiz. And number passed and failed quiz of students.



SMoodle
Statistic Moodle
logout [USER NAME]

Welcome to SMoodle

Assignment Statistis
Find out the list of assignments in the course, the number of attempts done by students for each assignment, the total time spent by the student on the assignment.

Quizzes Statistic
Find out the quizzes list with open/close time, student grade, the number attempts by the student, the time spent in the quiz. The total attempts in each quiz. And total numbers of passed and failed students.

Forum Statistic
Find out the forums list with the name of student joined the forum the number of post and/or view done by the individual in the forum. The total number of student joined the forums, the total number post and read to the forums.

Your Activity for the last 30 days
Your activities for the last 30 day's.
No activities for the last 30 day's.

Student Activities
Students activities for the last 30 day's

Courses List
Your courses for this current semester

OPERATING SYSTEM
Last Access: 2012-06-28
Assignment 2 | Quizzes 4 | Forums 2 | Students 8
Topics covered in this course include: Overview of operating systems, operation system principles, concurrency, scheduling and dispatch, memory management, real-time and embedded systems, and system performance evaluation.

SYSTEM ANALYSIS AND DESIGN
Last Access: 2012-06-28
Assignment 1 | Quizzes 1 | Forums 1 | Students 3
This course is about the development of computerized business system using CASE tools, 4GL and application generators, project management techniques, and back-end phases of SDLC. The Lab work is designed to develop and implement real business systems through lab projects.

Search for student's
Use student full name or email
Search

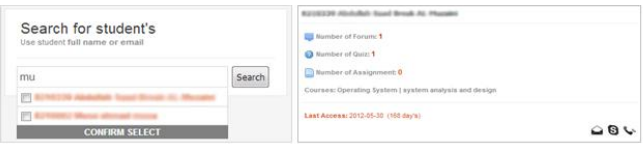
Fig. 5 SMoodle homepage snapshot

In forum block the educator can find out the list of forums available along with the name of students who joined the forum, the number of posts and reads done by the students in the forums.

Last but not least, the student block, with this block the educators can view the list of students who enrolled in his/her course along with the last time the students access the course. Student block also show the student's activities for a certain period for instance, the last 30 days. Student activities are determined by how much the students are seriously involved in particular course, for instance, the number of assignment submitted, the number of forums posts and reads, etc. See fig. 7 for student's page snapshot.

SMoodle home page gives more capabilities for the educators, see fig. 5. The list of courses currently taught by the teacher in the Moodle, along with the last time the course accessed by the teacher through the Moodle, and number of assignments, quizzes, forums, and students. Teacher activities and student's activities for last 30 days are presented using bar charts.

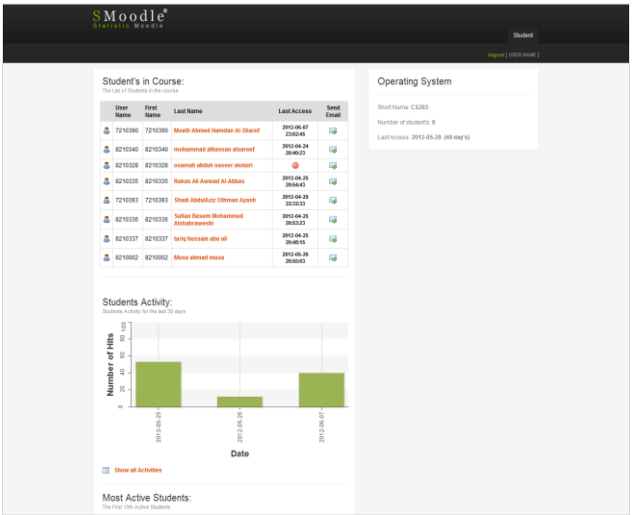
The educators can search for any particular student's already enrolled in the course using the search capability, see fig. 6 (left side). Students' full name or email or combination of both can be used as search criteria. The search results (right side), shows the basic information, such as, student full name, forums, quiz, assignment numbers, student enrolled courses, and the lass access time.



Search for student's
Use student full name or email
mu
Search
□ [redacted]
□ [redacted]
CONFIRM SELECT

[redacted]
Number of Forum: 1
Number of Quiz: 1
Number of Assignment: 0
Courses: Operating System | system analysis and design
Last Access: 2012-05-30 (168 day's)

Fig. 6 Student search box (left), search result page (right)



SMoodle
Statistic Moodle
Student
logout [USER NAME]

Student's in Course:
The List of Students in the course

User Name	First Name	Last Name	Last Access	Send Email
7210390	7210390	Moath Ahmed Hamdan Al-Sharef	2012-06-07 23:02:45	✉
8210340	8210340	mohammad alhassan alsareef	2012-04-24 20:40:23	✉
8210328	8210328	osamah abduh nasser alotairi	⛔	✉
8210335	8210335	Rakan Ali Awwad Al-Abbas	2012-04-25 20:54:43	✉
7210393	7210393	Shadi AbdulAziz Othman Ayash	2012-04-29 22:32:33	✉
8210336	8210336	Sultan Basem Mohammed Alshabraweshi	2012-04-25 20:53:23	✉
8210337	8210337	tariq hussain abu ali	2012-04-25 20:40:15	✉
8210002	8210002	Musa ahmad musa	2012-05-26 20:55:03	✉

Operating System
Short Name: CS203
Number of student's: 8
Last Access: 2012-05-28 (49 day's)

Students Activity:
Students Activity for the last 30 days

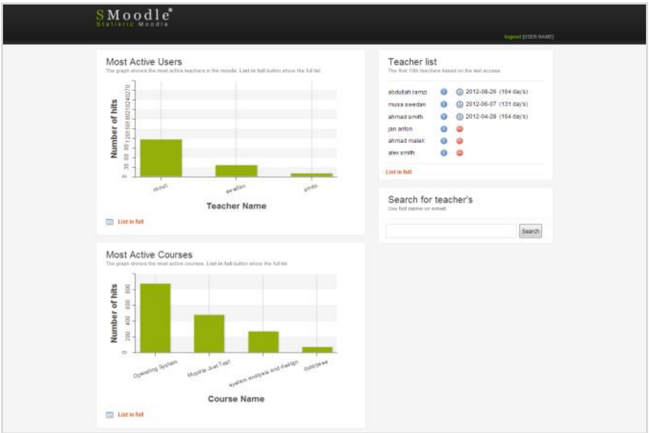
Show all Activities

Most Active Students:
The First 10th Active Students

Fig. 7 SMoodle student's page

B. *SMOODLE ADMINISTRATOR LEVEL*

The functions available for administrator level are different for the teacher level. It's mainly targeting the administrator and management person. The administrator home page (see fig. 8) shows useful functions to support the administrator decision regarding Moodle users, such as, the most active (i.e. they frequently upload course materials, create quiz and assignemnts, and so on) and non-active teachers in the Moodle system for a certain period, the most active courses in the whole Moodle system, the teachers list (Top-Right side) shows all Moodle educators, in ascending ordered by the last access time. As teacher level, searching capability is also available, but the search is applicable only on the teachers instead of students. Teacher full name and/or email are used as searching criteria.



abdullah ramzi 2012-08-26 (164 day's)
musa swedan 2012-06-07 (131 day's)
ahmad smith 2012-04-28 (164 day's)
jan anton
ahmad malek
alex smith

Fig. 8 Administrator home page

V. CONCLUSION

In this research, a 1883 students distributed on 22 courses of various departments in YU college are taken to study the students' learning behaviors patterns by using the dataMining algorithms, namely, classification (Decision Tree), clustering (K-Means) and association mining rule (FP-growth). Extracted knowledge's from these algorithms is used to support educators in the Moodle system to predict students' final outcome (use classification to predict if student will fail or pass the course according to a certain attributes, for instance, the number of quizzes, assignemnts or forums students involved in), and also to better understanding students' behaviors (use clustering to create a groups of students that share the same characteristics). Rapid Miner software is used to apply these algorithms. It is one of the most simple use and comprehensive data mining tools.

The Statistic Moodle or SMoodle developed with the purpose to support the educators in the electronic educational environment (i.e. the Moodle System) with a statistical tool that is more user-friendly interface than the Moodle system. SMoodle is web application written in PHP and Java Script languages.

REFERENCES

[1] Romero, C., Ventura, S. (2007). *Educational Data Mining: a Survey from 1995 to 2005*. Expert Systems with Applications, 33(1), 135-146.

[2] Klosgen, W., Zytkow,J. (2002). *Handbook data mining and knowledge discovery*. New York: Oxford University Press.

[3] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996). *From Data Mining to Knowledge Discovery in Databases*.

[4] Cristóbal Romero, Sebastián Ventura, Enrique García (2007). Data mining in course management systems: Moodle case study and tutorial.

[5] Baker, M.,(2010). *Data Mining for Education*. In McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition), vol. 7, pp. 112-118. Oxford, UK: Elsevie.

[6] Dougherty, J. Kohavi, M. Sahami, M. (1995). *Supervised and unsupervised discretization of continuous features*. In Int. Conf. Machine Learning Tahoe City, CA (pp.194–202).

[7] RapidMiner. (2012) rapid miner homepage. [Online]. Available: http://rapid-i.com

[8] Moodle. (2012) moodle homepage. [Online]. Available: http://www.moodle.org/

[9] Ext-JS 4 (2012) ext-js 4 homepage. [Online]. Available: http://docs.sencha.com/ext-js/4-0/

[10] Han Jiawei, Kamber Micheline, *Data Mining: Concept and Techniques*, 2nd edition. Morgan Kaufman, 2006.

[11] Liu, H. et. al:Discretization :A*n Enabling Technique.Data Mining and Knowledge Discovery*, 6,393-423.(Pub 2002).

[12] Jiawei Han , Micheeline Kamber.: *Data Mining Concepts and Techniques*, Morgan Kaufamann Pubisher(2009).