ГОДИШЕН ЗБОРНИК
2013
YEARBOOK
2013

ГОДИНА 2   **VOLUME II**

GOCE DELCEV UNIVERSITY - STIP
FACULTY OF COMPUTER SCIENCE

ГОДИШЕН ЗБОРНИК
2013
YEARBOOK
2013

СОДРЖИНА
CONTENT

ЕДУКАТИВНО ПОДАТОЧНО РУДАРЕЊЕ СО MOODLE 2.4

Зоран Милевски[1], Зоран Здравев[1]

[1] Факултет за информатика, Универзитет „Гоце Делчев" - Штип
milevskiz@gmail.com, zoran.zdravev@ugd.edu.mk

Апстракт

Околините за е-учење имаат за цел да обезбедат ефикасни методи за учење, да овозможат корисниците во кое било време да пристапат кон одредени ресурси, да постават решенија за одредени проблеми, да бидат оценети за нивниот труд и сл. Една од попознатите такви околини е системот за е-учење Moodle. Овие околини како Moodle користат и складираат големи количини на податоци, но во повеќето случаи не задоволуваат поголем дел од барањата за нивна примена и недоволно ја прикажуваат активноста на учесниците при учењето. Целта на овој труд е со помош на техники за податочно рударење како што се класификација, кластерирање, статистики и регресија, да се опише процесот на селекција и добивање на податоци од базата на податоци на Moodle и да се креира контролна табла – веб базирана апликација што ќе комуницира со системот за е-учење Moodle и ќе обезбедува неколку нивоа на пристап и тоа: менаџерско, администраторско, наставничко и корисничко ниво, и практично ќе прикажува обработени податоци и извештаи кои ќе го подобрат пристапот на евалуација на поголеми групи на учесници во процесот на учење. Со тоа директно се решава и проблемот на наставниците во поглед на нивната поддршка при работа со ваков тип на платформи и големи количини на податоци.

Клучни зборови: *далечинско учење, е-учење, едукативно податочно рударење, Moodle, едукативни контролни табли, извештаи во повеќе нивоа.*

EDUCATION DATA MINING WITH MOODLE 2.4

Zoran Milevski[1], Zoran Zdavev[1]

Faculty of computer science, Goce Delcev University, Stip, Macedonia
milevskiz@gmail.com, zoran.zdravev@ugd.edu.mk

Abstract

The goal of e-learning environments is to supply effective learning methods, to enable the users to approach certain resources at any time, to set solutions for certain problems, assessment for the work etc. One of the best known environments of this kind is e-

learning system Moodle. These environments like Moodle use and save large amount of data in their databases, but in most cases they don't offer enough information of the course participants and their activities in the system. The aim of this work is, by the use of data mining techniques such as classification, clustering, statistics and regression, to describe the process of selection and acquiring data from the Moodle database, and to create dashboard - web based application, that would communicate with the e-learning system Moodle and supply multilevel approach as: manager, administrator, teacher and user level; and practically will improve the approach to evaluation of larger groups of participants in the learning process. This will help teachers to evaluate web activity of the students, to get more objective feedback and find out more about how the students learn. Also this dashboard will directly solve the teachers problems in the terms of dealing with this kind of platforms and big amounts of data.

Keywords: *Distance Education, E-learning, Educational data mining, Moodle, Educational Dashboard, Multilevel reports.*

1. Introduction

Web-based educational systems and their usage has increased rapidly in the last few years. The impact on this trend comes from the fact that neither teachers nor students are limited any longer to be at the same time on the same location, and additionally these online education-based systems are independent from any hardware platforms [6]. The approach to these platforms is only through internet browser and thus the dependence on different operative systems and their demands is neutralized. These educational systems have been installed in many universities, and even individual teachers use them with a goal of setting certain resources that will be easily approachable for certain groups of people.

Moodle (Modular Object Oriented Developmental Learning Environment) as an educational system is well known and widely used because it is open code and also satisfies greater part of the needs for its use, and it is also simple to use both for the teachers and the students as course participants [1], [2], [3]. Moodle accumulates great amount of different information that is very important when analyzing the students conduct and represents a gold mine of educational data. Moodle stores all the data of the activities in which the students are involved. Moodle also keeps data of the participants profiles, their activity in different courses, their sent assignments etc.

The e-learning system Moodle is used in 232 countries in the world and at the moment there are 79429 active Moodle web sites, whereas in Macedonia there are 39 Moodle web sites most of which belong to the universities in the country [14].

Although Moodle, as well as the other systems of this type, offers tools for reports and view of the more important activities of the course participants, when it comes to a bigger number of students it becomes hard to follow their activity. On the other hand, although the goal of the e-learning systems is to motivate the

student by the use of multimedia materials to make studying more interesting and of higher quality, they don't always succeed in keeping the student's focus on the learning itself. Instead of learning they use the opportunities that the system offers in the direction of social communication between them (chat) [5]. Whatsoever, to make studying more effective, it is important to supply personalization of the contestants, based on their activity, an opportunity to analyze the participants in different courses, prediction of the results of the participants and better survey of the activities of the students. A promising area, when it comes to fulfilling this goal is data mining, and in this case it is educational data mining with the Moodle 2.4 database [3], [6].

Educational data mining means selective extraction of the kept data of large databases, their processing with the use of several educational techniques of data mining such as classification, clustering, statistics, regression etc. and acquiring the processed data that would improve the approach to larger groups of participants in the learning process [7]. The acquired information can be used not only by the teachers, but the students themselves too. They can get recommendations and directions for certain activities and resources that would improve their learning, where the teacher can get the feedback necessary for the evaluation of the students activity, separating the students in groups based on the need for their monitoring, finding the frequent mistakes made both by the students and the teachers; view into the activities assigned for the students; and have greater effect than the others [6].

Web based application connected to the active Moodle database can provide several levels of approach:
- Manager level approach,
- Administrator level approach,
- Teacher level approach and
- User level approach.

All of these roles are included in the, so called, dashboard and they will be explained in this study.

This dashboard is external application made to be very easy to use, because all needed reports for the various users are simplified in one place and enable a survey of some information kept in the new dashboard database, while the standard reports don't give such view with the use of standard reports.

2. Data analyses in the Moodle database

It is very important what kind of data is kept in the database and the more data is processed the more information can be acquired. Figure 1 illustrates the process of educational data mining and the way this process works [2].
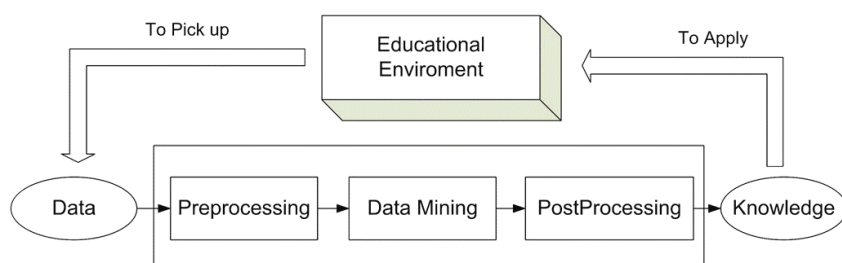
Слика 1. Процес на едукативно податочно рударење
Figure 1. Process of educatonal data mining

Educational data mining is an interactive process in which not only the processed data can be acquired, but it can also be filtered so that a certain decision can be made. The process consists of gathering information about the students' interaction within the process, than data processing so that they can be transformed into a relevant format to be mined. Data mining is applied, i.e. algorithms are used that provide and summarize the acquired interests about a certain user (teacher, student, manager etc.). Finally, the results are interpreted, evaluated and represented [2], [6].

Слика 2. Приказ на Moodle извештај
Figure 2 Moodle log report screen

Moodle stores every click of the user and its system navigation. Figure 2 shows a scheme of modest report record of Moodle about the site activities. Records can be filtered by course, participant, data and type of activity [15]. Teachers can use this report to follow the course participants activity, what they do and when. For activities such as quizzes the report contains data about the results, the time length of the quiz activity, as well as detailed analyses of every answer of the student. These reports are useful, but at the same time they are not clear enough. For a more effective view for the teacher, besides course activity it is important to be able to see which of the activities attracted greatest attention, which is the least visited material, in the quiz section, besides detailed analyses of every

answer, which is the question that was answered by the smallest number of students, analyses of the results of one student or all the students in several quizzes etc. [4], [13].

Moodle doesn't keep these records as text but keeps them in relational MySql database. Moodle database has around 145 related tables, but all of them are not necessary to implement educational data mining. Table 1 schemes the more important tables with their description that would be used for getting raw data so later be processed with several different techniques [6].

Табела 1. Поважни табели во базата на податоци на Moodle и нивни опис
Table 1. More important tables in Moodle database and their description

Table name	Table description
mdl_assign	Homework data
mdl_assign_submission	Sent homework status for a student
mdl_assign_grades	Homework score
mdl_choise, mdl_answers и mdl_option	View of questionnaire answers
mdl_course	Created courses data
mdl_enrol	Type of log on, course logging password
mdl_forum и mdl_forum_discussions	Created course forums and survey discussions started by a certain user
mdl_posts	Forum answers with date, user, message,
mdl_read	Forum participants activity
mdl_grades и mdl_grade_letters	Summarizing participant grades and score criteria and final grade
mdl_lesson	Set lessons data
mdl_log	Log for every student action
mdl_message и mdl_message_read	Survey of sent messages
mdl_question	Question base for the quizzes with answers
mdl_quiz	Quiz questions, answers
mdl_user	All users information
mdl_enrolments	Enrolled users in a certain course
mdl_user_lastaccess	participant's last course access

Data preprocessing provides data to be transformed in relevant format for data mining to be applied. Before using data mining it is important to identify the necessary user, the course he is enrolled etc [9], [10].

For example, to show the number of resources (figure 3) there is created new i.e. warehouse that is linked with the original Moodle database and it is filled with data in sertain time period.
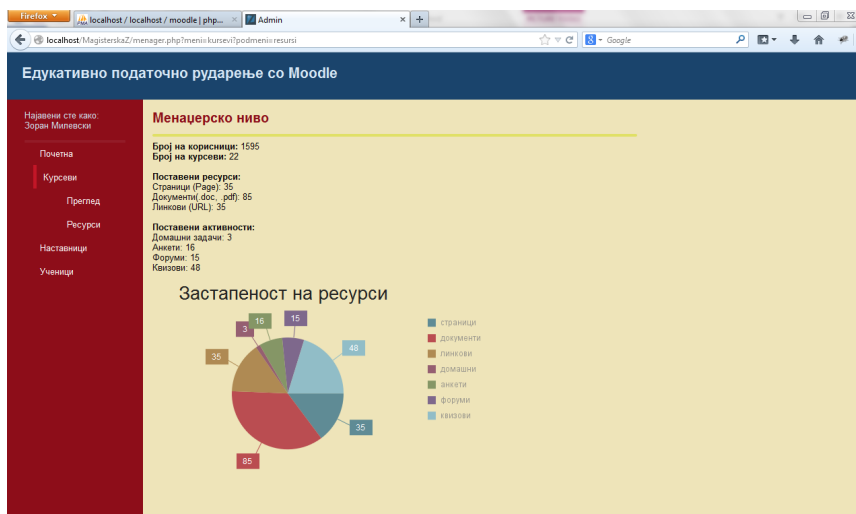
Слика 3. Страница за ресурси во контролната табла
Figure 3. Resources page in the dashboard

To get certain reports it is important to analyze several tables from the database so that a summary of the system activities can be provided, to get user-friendly results scheme. That is the aim of this research, with which we consider that system users (all within their own role) will get view in their activities.

3. Data mining of analyzed data

Besides analyzing the data in the Moodle database, it is very important how the data will be grouped in order to achieve the required effect.
For that purpose we hold up on data mining and we use some of the known techniques that can provide us all necessary information and data in the effort to give the teacher simplified view on the processed knowledge.

In e-learning systems clustering can be useful for finding similar characteristics students clusters, revealing the user conduct and grouping the students into several groups: students who are active in the system, discuss in forums, send homework, spend some time in the system in checking different contents etc. [6].

In this research we will divide students into three clusters as follows: cluster 0 (inactive), cluster 1 (very active), and cluster 2 (active course participants). Cluster 0 is characterized by students who haven't sent homework, have read only few messages, took only few quizzes and spent very little time in checking the resources, activities and forum participation. Cluster 1 is characterized by students who have sent at least one message in the forum, have read at least three messages, have passed successfully at least half of the quizzes and have finished less than half of them unsuccessfully and have high score and grades. Cluster 2 is characterized by students who have lower score than students in

cluster 1 and more than the students in cluster 0. In this way the teacher can use these information so that he can divide the students into groups of different type of students for example at least one student from cluster 1 and students from the other clusters or a group of students from cluster 1 who would work on problem assignments of higher degree than the others [6], [12].

Classification of participants is used to discover potential students with similar characteristics for a definite specific pedagogical strategy, to predict the final results for a group of students, even to identify the students who need motivation to get better results.

We divide the students into bad, good and excellent by generating decision trees that involve certain classification rules. Our goal is to classify students in different groups depending on their activity in Moodle. Table 2 represents the knowledge by decision tree with if-else rules. This process goes on until all data are classified perfectly or we run out of attributes. Students with lower number of passed quizzes are classified as weak students, students with bigger number of quizzes are classified as excellent and the students with an average number of quizzes as good and of course taking into account the total time spent on resources and activities, the number of sent homework assignments etc. [6], [9], [11].

Табела 2. Множество правила генерирани од одлучувачко дрво
Table 2. Rule set generated by Decision Tree

```
if(n_quiz=low) then mark=bad
else if(n_quiz=medium) then {
    if(total_time=low) then {
        if(view_resource=low) then mark=bad
            else if view_resource =medium) then {
                if(forum_post=low) then mark=bad
                    else if(view_resource=medium) then {
                        if(total_assigments=high) then
mark=good
                                else if(overall_core=high) then
mark=excellent
}
    else if(total_time =medium) then {
        if(view_resource=low) then mark=bad
            else if(view_resource=medium) then mark=good
                else if(view_resource=high) then
mark=excellent
                        if(overall_score=good) then
                            if(forum_post==good) then
                                mark=excellent
        }
```

Teachers can use this information from these rules to get an overview of the course activity and the classification of the course participants. For example, it is obvious that the main discriminator in this case are the successfully realized

quizzes, but there are also other decisive factors that would help the teacher to decide about the type of activities he would use in the future, to decide which activities not to use in the future due to the bad results or their insufficient attractiveness among the other activities and resources. The teacher can and decide which of the students have difficulties in learning, which topics are more difficult to overcome, so that he can react on time.

The last mining technigue that will be described in this paper is regresion. It is the easiest technique to use, but is also probably the least powerful. Regresion is a data mining function that predicts a number. A regression task begins with a data set in which the target values are known. In the model build (training) process, a regression algorithm estimates the value of the target as a function of the predictors for each case in the build data. These relationships between predictors and target are summarized in a model, which can then be applied to a different data set in which the target values are unknown [19]. For example, in our case study, using regresion we can predict the students grades, based on observed data for many students activities over a period of time.

For our dashboard, different roles of users (manager, administrator, teacher, user), display different reports gained with the mining techniques.

a. Manager level acquired data

Manager level enables manager role data survey and getting reports that enable to follow the activity of all the participants in the system, number of courses, set materials and resources, realized quizzes etc.

After logging as role manager, the home page (figure 4) gives the manager quick view of the activity of the users and review of the number of the participants in the courses.
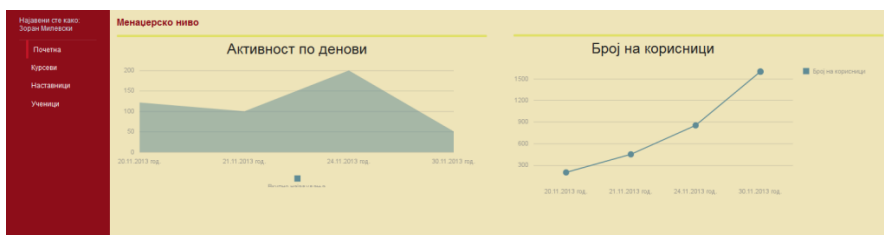
Слика 4. Почетен поглед по најавување во улога на менаџер
Figure 4. First view after login in the role manager

The menager can look over all system registered users, users that have confirmed their registration and those that haven't confirmed the registration, whereas in the submenu user activities he can see the last system access of the users, shown for a certain date. An example for how much time the participants (teachers/students) spent in the system is given at figure 5.

Табеларен приказ на корисници со улога на наставници				
Прикажи 10 ▼ записи				Пребарување:
Име ◆ Презиме ◆	Број на курсеви ◆	Зачленет ◆	Последно најавување ◆	Вкупно поминато време ▼
Здр....... ....вски	4	02.09.2012 год.	28.09.2013 год. 18:21 часот	6 дена 16 ч 12 мин 33 сек
Бла...............	2	02.09.2012 год.	25.09.2013 год. 07:47 часот	21 ч 37 мин 38 сек
Vele................ ....га..	2	13.11.2012 год.	13.03.2013 год. 11:58 часот	13 ч 21 мин 40 сек
Ан...........................	2	17.09.2012 год.	01.12.2012 год. 22:19 часот	06 ч 55 мин 28 сек
Кат.......... .... М..........	2	15.10.2012 год.	28.09.2013 год. 19:07 часот	04 ч 04 мин 30 сек
Ан.............. Костовски	4	06.12.2012 год.	27.09.2013 год. 07:29 часот	03 ч 02 мин 13 сек
Е....................	1	05.12.2012 год.	05.12.2012 год. 11:10 часот	01 ч 45 мин 27 сек
В.................	1	25.03.2013 год.	25.03.2013 год. 21:18 часот	01 ч 16 мин 44 сек
М..................	1	02.04.2013 год.	24.04.2012 год. 19:44 часот	01 ч 06 мин 01 сек
Кле......... .... Петре....	1	02.09.2012 год.	25.10.2012 год. 10:24 часот	00 ч 53 мин 36 сек
Прикажување 1 до 10 од вкупно 11 записи			Прв Претходен 1 2 Следен Последен	

Слика 5. Поминато време во системот од извештајот за корисничка
активност
Figure 5. Time spent in the system from the user activity report

The course menu schemes a list of courses in categories and a total number of, list of all courses and an opportunity to choose a certain course. After a course is chosen a view of the number of discussions without content display is acquired, but a view of the activities of the discussion participants, number of started topics, number of theme answers. A list of all resources and access to all of them with additional details for most visited and least visited resource. A survey of all the activities such as homework and its assessment with a list of students with highest and lowest score is also available. This part also offers a survey of all questionnaires and their results as well as a possibility to print the questionnaire results.

User portfolio enables individual course users' data preview, as well as table preview for all the course participants by viewing the activity (inactive, active, very active), division in categories according to the assessment into bad, good and excellent, prediction of whether they will complete the course successfully or not and complete summary.

The user portfolio view gives contrastive analysis of a student's results in several courses and all the activities that characterize the student.

b. Administrator level acquired data

The administrator is the user of the application who has all the privileges. The application administrator has the overall view of the user from the manager level. There is an additional opportunity to give tasks in precisely defined period of time (when it is expected not to have any activities in the system) to keep the data in the new database, as well as to archive the previous preview, because previously all overviews through cancel procedures are read from the Moodle database that is linked to the new database, the warehouse. Besides that the administrator has the right to register new users of the system from external bases for example excel documents.

c. Teacher level acquired data

The teacher has a similar role in the application to the manager, with the little difference that the manager has an overview of all courses and all users of the application, whereas the user with the teacher's role can view only the data that refers to the courses that he has created and the system users that are participants only in his courses.

Additionally the teacher has better view of the section that refers to the results of the student's homework, quizzes and their activity in the system.

The teacher can see which of the questions the participants have been answered correctly and which not so that he can direct the participants to find out the correct answers in the following lessons. For example, in the quiz section, to be able to see the results according to the standard Moodle report several steps are required in order to get a table view, so the application enables getting view with only one click, automatic sorting of the results and a percentage representation of the score as well as another column with a grade in form of letter.

In the section that refers to the comparison of the results and the activity of the students in the other quizzes, the teacher can see only the analyses of the activities and the processed data only for the courses that he has created, but not for all the other quizzes in the e-learning system.

d. User level acquired data

User with the user level role in the application is in fact the student who participates in one or several courses of the e-learning system Moodle. The user has a username and a password as in the profile he has created on the e-learning system Moodle itself. The data that the user can see is from the user portfolio of the manager and the teacher and refer only to the logged user. In this way the user can view in which subject he participates, to see his activity, results, comparison of the activities in different courses etc.

Besides these views the user gets certain suggestions by the teacher for the necessity to pay more attention and to be more active in the working obligations within the course in order to motivate him to accomplish better final results.

4. Conclusion

This work gives analyses of the data from the database in the e-learning system Moodle and gives survey of the results from the data mining with the use of several techniques applied on the application that offers several levels of approach. It is necessary to integrate the data mining tools in the e-learning environments which is the goal of this research, because in this way all these data mining techniques will be applied in a single application and the feedback and the acquired results will be directly applied on the e-learning environments [16].

Here are several data mining techniques that can be used for acquiring processed results and reports in the process of learning, and they are not complicated to be used by the teachers. That is why this approach of creating

dashboard – web based application, which is user-friendly and give better control when it comes to larger groups of students when the standard reports reduce the control clarity and the ability to evaluate their results at the end of the course [3], [5], [16].

Nowadays, data mining tools are too complex to be used by the educators and their futures go beyond the scope of what educator might to do. By creating a dashboard that would communicate with the e-learning system Moodle, the teachers can easily evaluate web activity in order to get more objective feedback, and find out more about students capability in successfully passing the exam. Also this dasboard will directly solve the teachers problems in supplying support in dealing with various kind of algorithms. It could also be oriented towards the academics and administrators responsible in order to obtain parameters about how to improve site efficiency and adapt it to the behavior of their users, have measures about how to better organize institutional resources (human and material) and their educational offer, enhance educational program offers, etc.

Литература (References)

[1] Ramaswami M., and Bhaskaran R.: A Study on Feature Selection Techniques in Educational Data Mining”, vol.1, Journal Of Computing, ISSN:2151-9617, https://sites.google.com/site/journalofcomputing (December 2009)

[2] Elatia S., Ipperciel D., Hammad A.:Implications and Challenges to Using Data Mining in Educational Research in the Canadian Context, Canadian journal Of Education, pp. 101--119 (2012)

[3] Baradwaj B. K., Pal S.:Mining Educational Data to Analyze Students' Performance, (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 2, no. 6 (2011)

[4] Yadav S. K., Bharadwaj B., Pal S.:Mining Education Data to Predict Student's Retention - A comparative Study, (IJCSIS) International Journal of Computer Science and Information Security, vol. 10, no. 2 (2012)

[5] Cocea M., Weibelzahl S.: Disengagement Detection in Online Learning - Validation Studies and Perspectives, IEEE transactions on learnin technologies, vol. 4, no. 2 (April-June 2011)

[6] Romero C., Ventura S., García E.:Data mining in course management systems - Moodle case study and tutorial

[7] BAKER R.S.J.D., YACEF K.:The State of Educational Data Mining in 2009 - Review and Future Visions

[8] Retalis S., Papasalouros A., Psaromiligkos Y., Siscos S., Kargidis T.: Towards Networked Learning Analytics – A concept and a tool

[9] Romero C., Ventura S., Espejo P. G. and Hervás C.: Data Mining Algorithms to Classify Students, The 1st International Conference on

Educational Data Mining, Montréal, Québec, Canada, pp. 8-18 (June 20-21, 2008)

[10] Yadav S. K., Pal S.: Data Mining - A Prediction for Performance Improvement of Engineering Students using Classification, World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 vol. 2, no. 2, pp. 51-56 (2012)

[11] Dash M., Liu H.:Feature Selection for Classification, An International Journal of Intelligent Data Analysis, vol. 1, no. 3,2006, pp. 131-156 (1997)

[12] Chen G., Liu C., Ou K., Liu B.: Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology, Journal of Educational Computing Research, pp. 305–332 (2000)

[13] Anozie N., Junker B.W.: Predicting end-of-year accountability assessment scores from monthly student records in an online tutoring system, Educational Data Mining AAAI Workshop, pp. 1-6, California, USA (2006)

[14] Moodle org. LMS Moodle official site. web. 11 Apr. 2013, http://moodle.org

[15] High school Dobri Daskalov, E-learning Moodle, Kavadarci, R. of Macedonia: n.p., 2009. web. 11 Apr. 2013, http://moodle.dobridaskalov.edu.mk

[16] Darrell M. W.:Big Data for Education - Data Mining, Data Analytics, and Web Dashboards, U.S. Department of Education Office of Educational Technology, Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics", pp. 36 (2012)

[17] Oracle. Oracle Data Mining Concepts. Release 1 (11.1), Oracle Data Mining Concepts , web.7 Sept. 2013