



## Data Mining in Learning Management Systems for Predicting Student's Performance: Moodle and Rapidminer Case Study

Arti Tyagi, Manisha Jailia, Manisha Aggarwal

AIM & ACT,  
Banasthali University, India

Ashish Tyagi

Electronics & Communication,  
Suresh Gyan Vihar University, India

**Abstract**— Educational Data mining nowadays became an emerging field used to extract knowledge and discover patterns from E-learning systems databases. Educational data mining is mainly concerns with developing methods and techniques for discovering knowledge from data that come from educational domain. The increasing popularity and the use of e-learning systems increased through the last decades. Students or learner's produce a lot of data through their use of and interactions with the e-learning system, that data is often not exploited. Since there is a vast amount of such data which is unexplored and can be used for many purposes. For exploring data, data mining techniques were used as the appropriate solution. This work is a specific application of data mining in learning management systems and a case study of Moodle and Rapidminer. Our objective is to introduce it in both a theoretical and practical way to all users interested in it. In this paper we show practical experience with specific e-learning system and applied data mining technique for evaluating the student's performance and group them according to their performance.

**Keywords**— Educational data mining, Learning management system, Classification, Data mining, Decision tree.

### I. INTRODUCTION

#### BACKGROUND

Data mining or knowledge discovery in databases (KDD) is the automatic extraction of implicit, unexplored and interesting patterns from large data collections that is databases. Data mining is a multidisciplinary area in which several computing paradigms converge: decision tree construction, rule induction, artificial neural networks, instance-based learning, Bayesian learning, logic programming, statistical algorithms, etc. And some of the most useful data mining tasks and methods are: statistics, visualization, clustering, classification, association rule mining, sequential pattern mining, text mining, etc [1].

The knowledge discovered can be used not only by providers (educators or teachers) but also by the users themselves (students or learners), so it can be oriented for different ends from each particular point of view. It could be oriented towards students in order to recommend learners' activities, resources, suggest path pruning and shortening or simply links that would favor and improve their learning, or oriented towards educators in order to get more objective feedback for instruction, evaluate the structure of course content and its effectiveness in the learning process, classify learners in groups based on their needs for guidance and monitoring, find learner's regular as well as irregular patterns, find the most frequently made mistakes, find activities that are more effective, etc. 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 (optimal server size, network traffic distribution, etc.), have measures about how to better organize institutional resources (human and material) and their educational offer, enhance educational program offers, etc[1]. The application of various data mining techniques, tools and concepts in e-Learning systems will help to support educators or teachers to improve their e-Learning environment. Data mining techniques have also been addressed as complementary systems to LMS, and in particular to Moodle. This paper is oriented to the specific application of data mining in online or e-learning educational systems. It contains the general process of applying data mining to e-learning data, especially to Moodle usage information, the preprocessing step necessary for adapting the data to the appropriate format, the application of the main data mining techniques in e-learning and an example case study with Moodle data. This work describes the application of data mining techniques to the usage data of the Moodle course management system. Finally, the conclusion is outlined.

#### A THEORY FOR E-LEARNING SYSTEMS

The use of online or e-learning systems has grown exponentially in the last few years, spurred by the fact that neither students nor teachers are bound to a specific location. Such e-learning systems are sometimes also known as a Course Management System (CMS), Learning Management System (LMS), Learning Content Management System (LCMS), Managed Learning Environment (MLE), Learning Support System (LSS) or Learning Platform. These systems can offer a great variety of channels and workspaces to its users so that they are able to facilitate information sharing and communication between participants in a course, to let educators distribute information to students, produce content material, prepare assignments and tests, engage in discussions, manage distance classes and enable collaborative learning

with forums, chats, file storage areas, news services, etc. Some examples of commercial systems are Blackboard, WebCT, TopClass etc. and some examples of free systems are Moodle, Ilias, Caroline, etc. Nowadays, one of the most commonly used is Moodle (Modular Object Oriented Developmental Learning Environment) which is a free learning management system that enables the creation of powerful, flexible and engaging online courses and experiences [1].

These e-learning systems accumulate a vast amount of information which is very valuable for analyzing students' behavior and could create a gold mine of educational data. Learning management systems accumulate a great deal of log data about students' activities. They can record whatever student activities are involved, such as reading, writing, taking tests, performing various tasks, and even communicating with peers. They normally also provide a database that stores all the system's information: personal information about the users (profile), academic results, user's interaction data, etc. However, due to the vast quantities of data these systems can generate daily, it is very difficult to manage manually, and authors demand tools which assist them in this task, preferably on a continuous basis. Although some platforms offer some reporting tools, when there are a great number of students, it becomes hard for a tutor to extract useful information. They do not provide specific tools which allow educators to thoroughly track and assess all the activities performed by their learners and to evaluate the structure and contents of the course and its effectiveness in the learning process. A very promising area for attaining this objective is the use of data mining [1].

## **II. DATA MINING TOOLS FOR E-LEARNING SYSTEMS**

Several data mining tools which assist educator to extract knowledge and discover patterns can be used. Such tools can be generic or specific, commercial or open source. Some Examples of commercial data mining tools are: DBMiner, IBM SPSS Modeler, DB2 and Intelligent Miner etc. Examples of open source tools are Rapid Miner, Weka, Keel etc. In order to offer a better and more flexible service to educators there have been developed several specific educational data mining tools: Tools for association and classification and text mining – The Mining tool, EPRules, Simulog, Sequential Mining tool, O3R and KAON, MultiStar and CIECoF; Tools for statistics and visualization-Synergo/ColAT, GISMO, Listen tool TADAEd etc.[2]

## **III. PROCESS OF DATA MINING IN E-LEARNING SYSTEMS**

The development of online learning courses is a difficult task few years earlier in which the developer has to choose the contents that will be shown, decide on the structure of the contents and also to determine the most appropriate content elements for each type of potential user of the course. Due to these difficult tasks of taking these decisions, a one-shot design is hardly feasible, even when it is carefully done. It is required in most cases to evaluate and possibly modify the contents, structure and navigation of students' usage information. To simplify these, we need data analysis methods and tools to observe students' behavior and to assist teachers in detecting possible errors, shortcomings and possible improvements. The e-learning data mining process consists of the same four steps in the general data mining process as follows: first is collection of data (In this paper we are going to use the students' usage data of the Moodle system), second step is preprocessing of the data (In order to preprocess the Moodle data, we can use a database administrator tool or some specific preprocessing tool), third step is application of data mining technique (In order to do so, we can use a general or a specific data mining tool) and the last step is interpretation and evaluation of the results - The results or model obtained are interpreted and used by the teacher for further actions. The teacher can use the information discovered to make decisions about the students and the Moodle activities of the course in order to improve the students' learning.

## **IV. PREPROCESSING MOODLE DATA**

The word Moodle is an acronym for Modular Object-Oriented Dynamic Learning Environment. Its modular design makes it easy to create/add/update/remove courses, adding content that will engage learners in the learning activities. Moodle LMS offers several reports on the students' activities. Moodle contains logs of all activities that students perform. Logging is record keeping that contain track of which materials students have used. Moodle keep track of every click that students make for navigational purposes and has a modest log viewing system built into it. Log files that moodle contains can be filtered by course, participant, day and activity. The administrator or teacher can use these logs to determine who has been active or use the course, what they did and when they did it. For activities such as quizzes, not only the score and elapsed time are available, but also a detailed analysis of each student's performance is available. Teachers are able to get full reports of the activities of each student, or of all students for a specific activity.

Data preprocessing allows the original data to be transformed into a suitable shape to be used by a particular data mining algorithm or framework. So, before applying a data mining algorithm, a number of general data preprocessing tasks can be done.

### ***Mining Moodle Logs***

Data mining done on based on Moodle a log file tries to discover usage patterns by making use of data mining techniques with the purpose of understanding and better serving the user and the application itself [2]. The summarization table created from the Moodle logs tables such as: mdl\_user, mdl\_user\_students, mdl\_log, mdl\_quiz, mdl\_grades, mdl\_final\_grades, mdl\_course, etc. It is required to create a new table in the Moodle database that can summarize the information at the required level (e.g. student). Student interaction data are spread over several tables. We have to create a new summary table which integrates the most important information for our objective. This table has a summary per row about all the activities done by each student in the course and the final mark obtained by the student in

the quiz and assignment. In order to create this table it is necessary to make several queries to the database in order to obtain the information about suitable students (*userid* value from *mdl\_user\_students* table) and courses attempted (id value of the *mdl\_course* table). The *mdl\_summarization* we have created have the attributes – *quizattempted*(name of the quiz attempted), *marks*(marks obtained in all quizzes) and *performance*(performance of the students on the basis of the marks obtained in all the quizzes).

### V. APPLYING DATA MINING TECHNIQUE TO MOODLE DATA

The purpose of data preprocessing of moodle data was to understand and prepare the required data and apply data mining techniques that allow teachers to detect patterns of use for further evaluation. This work is based on log analysis of Moodle LMS by applying data mining techniques to student's usage of data in order to evaluate their performance and to know about how the students learn on the E-learning environment. Data mining techniques are used to build relationships among Moodle LMS system components. Rapid Miner data mining tools are used to extract knowledge from the Moodle system [2]. In our work we are using the Rapid Miner data mining tool for applying data mining on the moodle data. Rapid Miner provides data mining and machine learning procedures including: data loading and transformation, data preprocessing and visualization, modeling, evaluation, and deployment. Rapid Miner is based on modular operator concept which facilitates rapid prototyping of data mining processes by way of nesting operator chains and using complex operator trees. Owing to the modular operator concept, the data mining processes are optimized because, by substituting or replacing one particular operator at a time and leaving rest of the data mining process design untouched, its performance can be evaluated. In our work we use the classification technique of data mining to mine moodle data. Moodle data is mined using tree induction method of classification technique for evaluating the student performance.

### VI. RESULT

In our work we are using the decision tree algorithm of tree induction in Rapid Miner data mining tool. We apply the decision tree algorithm on the *mdl\_summarization* table to obtain the desired results.

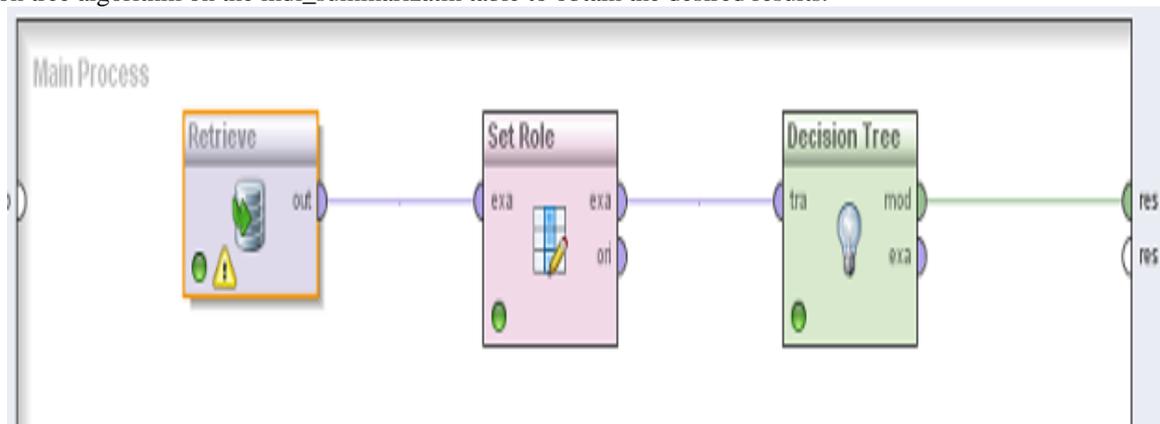


Figure 1. Showing the application of data mining operator

We change the decision tree operator parameter criterion settings different each time for obtaining different aspects of the same result or more detailed version of the same result. In the figure 2, we show the parameters of the decision tree tool of Rapid Miner .

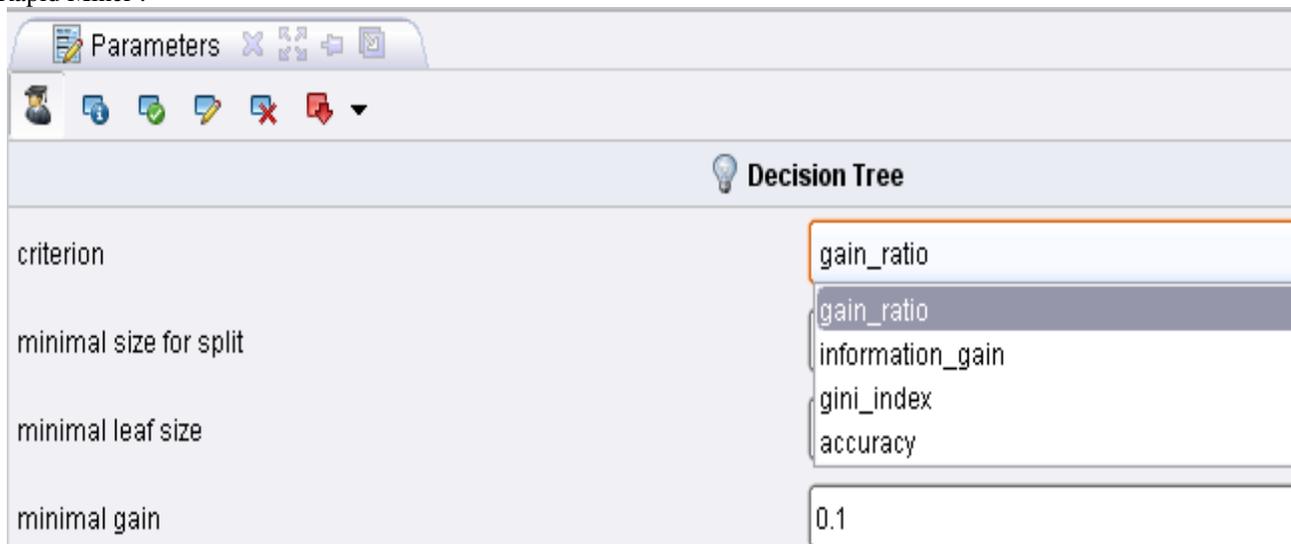


Figure 2. Showing the parameter settings of the decision tree operator

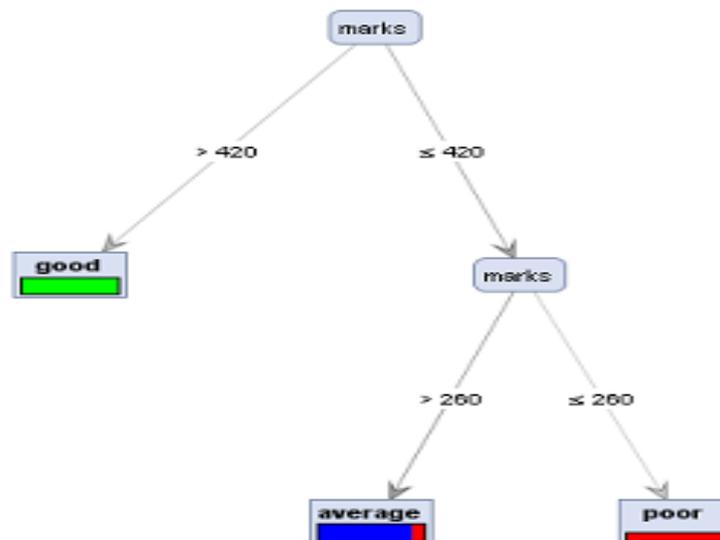


Figure 3. Showing the result of data mining

The figure 3, shows a set of IF-THEN-ELSE rules was generated from the decision tree operator, showing interesting information about the supervised classification of the students. The rules generated from the Tree Induction Algorithm operator were shown in figure 4.

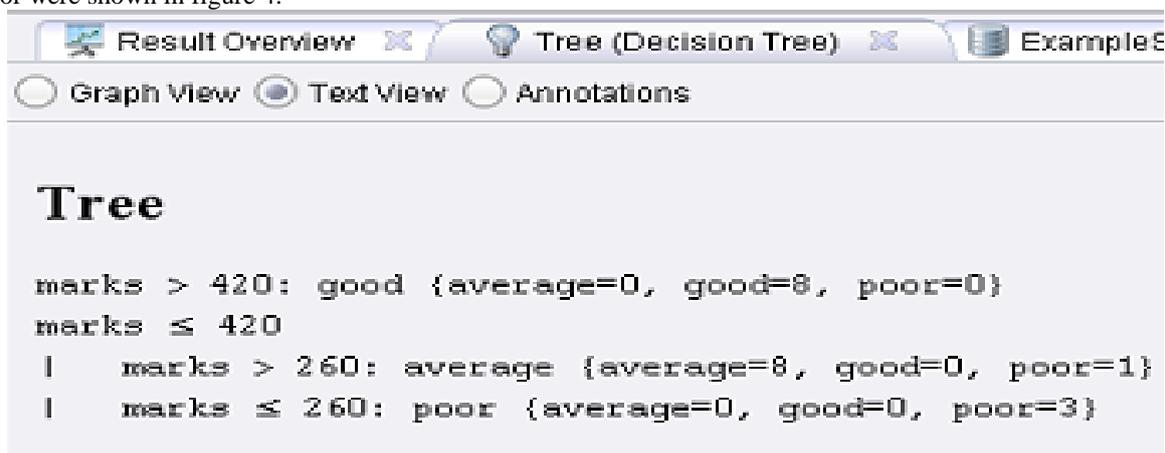


Figure 4. Showing the result of Tree Induction algorithm operator

The figure 5, shows the result of applying decision tree algorithm on the mdl\_summarization table but by changing the criterion settings of decision tree operator from gain\_ratio to information gain.

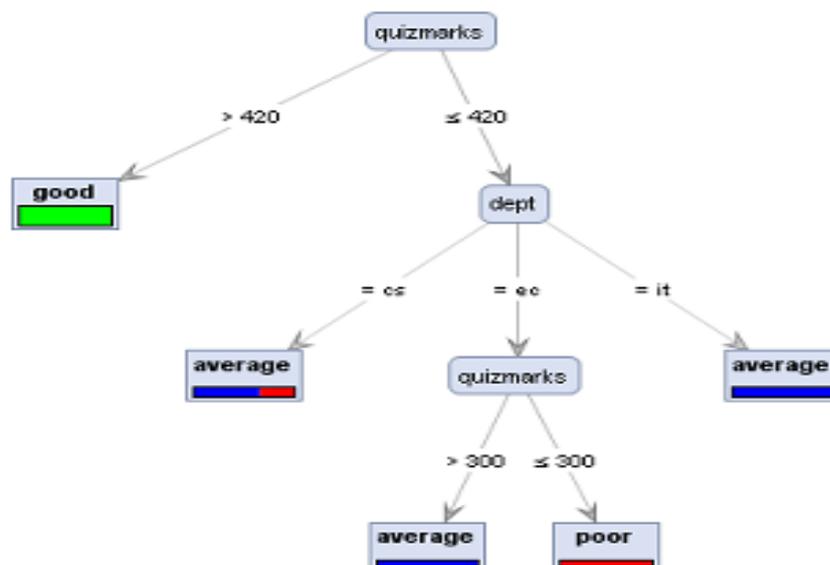


Figure 5. Showing the result of data mining

The above result is shown in the text view in the below figure, which also gives the information about the set of IF-THEN-ELSE rules was generated from the decision tree operator, showing important information about the supervised classification of the students

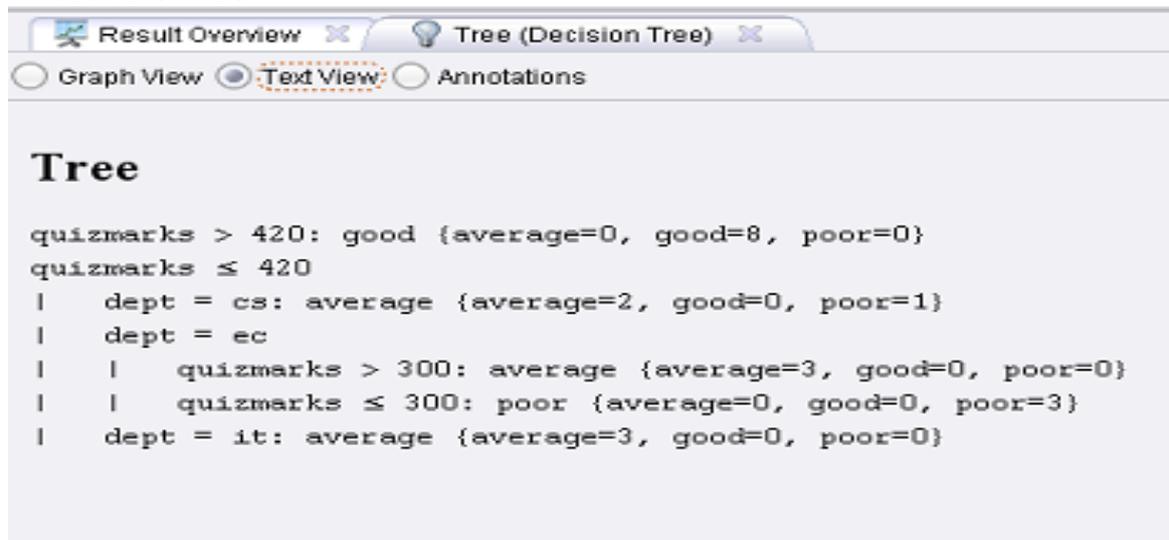


Figure 6. Showing the result of Tree Induction algorithm operator

The figure 7, shows the result of applying decision tree algorithm on the mdl\_summarization table but by changing the criterion settings of decision tree operator from information gain to accuracy.

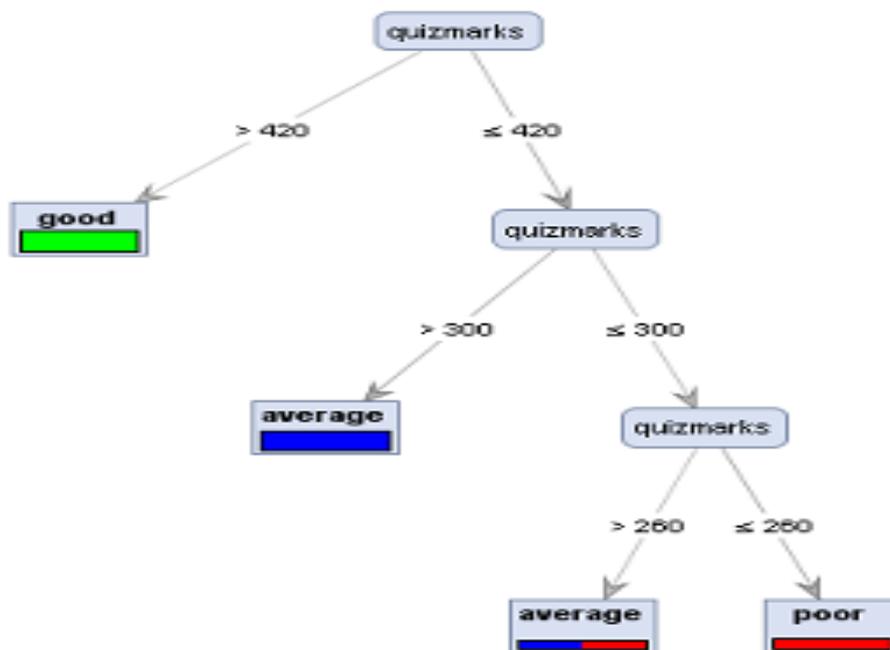


Figure 7. Showing the result of data mining

## VII. CONCLUSION

In this work we have done the evaluation of student's performance and divide them into three categories on the basis of their marks obtained in the quizzes they have taken. On the basis of this information, teachers are able to know about the student's or learner's performance, that how much student's have poor, average and good performance and decisions are taken accordingly to improve student's performance or improving content's quality.

## REFERENCES

- [1] Cristobal Romero, Sebastian Ventura and Enrique Garcia, "Data Mining in course management system : Moodle case study and tutorial".
- [2] Anduela Lile, "Analyzing E-Learning Systems Using Educational Data Mining Techniques" ,Mediterranean Journal of Social Sciences , vol. 2(3), September 2011
- [3] Divna Krpan and Slavomir Stankov, "Educational Data Mining for Grouping Students in E-learning System", Int. Conf. on Information Technology Interfaces, June 25-28,2012

- [4] Samrat Singh and Dr. Vikesh Kumar, "Classification of Student's data using Data Mining Techniques for Training and Placement Department in Technical Education", International Journal of Computer Science and Network(IJCSN), Volume 1, Issue 4, August 2012, ISSN 2277-5420.
- [5] Surjeet Kumar Yadav and Saurabh Pal, "Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification", World of Computer Science and Information Technology Journal(WCSIT), Vol.2, No. 2, 51-56, 2012.
- [6] Divna Krpan and Slavomir Stankov, "Educational Data Mining for Grouping Students in E-learning System", Int. Conf. on Information Technology Interfaces, June 25-28, 2012
- [7] Arne Duncan and Karen Cator, "Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics", U.S Department of Education, 2012.