Data Mining: A Prediction for Performance Improvement in Online Learning Systems

Manisha Jailia , Arti Tyagi
AIM & ACT, Banasthali University
Rajasthan , India

Abstract— *Data mining in education slowly evolves from the need of simple discovery of frequent patterns and regularities in large data sets toward interactive, user-oriented, on-demand decision supporting. Since data to be mined is usually located in a educational database, there is a promising idea of integrating data mining methods into database management systems (DBMS). The popularity and the use of online learning systems increased through last decades .Student's produce a lot of data through their interactions with the online learning system produce a lot of data, which is often not exploited .In this paper, we show how using data mining techniques can help discovering pedagogically relevant knowledge contained in databases obtained from Web-based educational systems or Online Learning Systems. These findings can be used both to help teachers while managing their class, understand their student's learning and provide proactive feedback to learners and teachers. This paper is a work of the specific application of data mining technique (Classification) in online learning management systems .*

Keywords— Online Learning System, Data Mining, Classification, Educational Database, Database Management System.

I. INTRODUCTION

"Data Mining-The non trivial extraction of implicit, previously unknown, and potentially useful information from data".

Data mining, *the extraction of hidden predictive information from large databases*, is a powerful new technology with great potential to predict future trends and behaviours allowing making proactive, knowledge-driven decisions. Data mining, also referred to as database mining or knowledge discovery in databases (KDD) is a new research area that aims at the discovery of useful information from large datasets [1]. Data mining uses statistical analysis and inference to extract interesting trends and events, create useful reports, support decision making etc. It exploits the massive amounts of data to achieve business, educational, operational or scientific goals. Data mining encompasses different algorithms that are diverse in their methods and aims .It also comprises data exploration and visualisation to present results in a convenient way to users. The field of Data Mining is concerned with finding new patterns in large amounts of data. Here we are interested in mining student models in a pedagogical perspective. The goal of our project is to define how to make data possible to mine, to identify which data mining techniques are useful and understand how to discover and present patterns that are pedagogically interesting both for learners and teachers.

Data mining provides a set of techniques which can help educational system to overcome several issues such as identifying students need, personalization of training and predicting quality of student interactions by analyzing student's trends and behaviors towards education in order to improve learning experience of students [2]. The application of data mining in e-learning systems is an iterative cycle in which the mined knowledge should enter the loop of the system and guide, facilitate and enhance learning as a whole, not only turning data into knowledge, but also filtering mined knowledge for decision making.

Examination plays a vital role in any student's life .The marks obtained by the student in the examination decide his/her future .Therefore it becomes essential to predict whether the student will pass or fail in the examination. If the prediction says that student tends to fail or obtained less in the examination prior to the examination than extra efforts can be taken to improve his/her performance[2].It is important to both the student and the faculty to know about the student performance ,student by knowing his/her weak points and strengths work on them to improve his/her performance and teachers by knowing student's performance know his/her knowledge state and make changes in his/her instructions.

Teachers Face A Dilemma they are being asked to assess students more, but every minute of class time spent in testing is a minute of instructional time that is lost. Online Learning Systems solves this dilemma, by letting students get individual feedback while their learning progress is assessed .Online Learning Systems is a free online platform that allows teachers to write and select questions, students get immediate and useful tutoring and teachers receive feedback on the performance of each individual student and of the class as a whole and adjust their instructional actions to influence student learning .By examining the feedback data, instructors can spot students who may need additional help or

encouragement to spend more time on the content and identify areas where the class as a whole is struggling[3]. An Online Learning System can be of any form like educational material which is readily available for distribution on the Web or privately over an internal network .The use of computers in learning and teaching advanced a lot through last decades through different online learning systems. The use of networked internet and intranet systems to disseminate training and educational information to a group of users .Student's learn ,explore content and by doing so they leave trail of log information. Online learning systems have the ability to capture streams of fine grained learner behaviours and the tools and techniques that can operate on the data to provide a variety of stakeholders with feedback to improve teaching, learning and educational decision making[4].Online learning system provides data that is transformed into information used by teachers in order to improve teaching process. Student's also benefit from that information by getting appropriate content and also by adapting their learning process .Educational data mining might can also be used for student modeling and online learning system evaluation.

II.   RELATED WORK

http://www.assistments.org. The ASSISTments online platform helps teachers write questions for assessments and then see reports on how their students performed. Students can get immediate tutoring while they are being assessed .**ASSISTments is an online tutoring program that** gives immediate feedback to teachers, students, school administrators, and parents .It has flexible content for teachers to use pre-built problem sets, edit pre-built problem sets, or build their own problem sets .It gives student problem solving strategy sequences ,knowledge state and strategy ,receive hints and tutoring to the extent they need .It uses information how student's respond to the problems and how much support they want from system to generate correct responses as assessment information .The teacher gets immediate feedback on how the student and the class performed on the problem set .This assists the teacher in making data drive adjustments in their lesson plans .The most important aspects of ASSISTments students get feedback while teachers get reports[4]

http://wayangoutpost.com/. Wayang Outpost is an intelligent tutoring system that helps middle and high school students study for standardized tests and adjusts instruction as they progress. Wayang Outpost and The Wayang Mathematics Tutor (Wayang Math) are machine tutoring systems that use interactive multimedia to help students improve math skills. Wayang Math is appropriate for middle and high-school students and is proven to increase scores for standardized math tests, such as the SAT, MCAS and CA-Star. It's also good for students of all ages preparing for college-level mathematics classes, and anyone who want to review math essentials. Wayang Math also helps teachers determine their students need areas and strengths .Tutoring sessions are customized to support lesson plans .As the student progresses through the math problem presented ,Wayang Outpost adjusts instruction ,presenting customized strategies for each student[4].

http://www.dreambox.com/.DreamBox is a highly adaptive learning math program for students . It is best used by individual students on their own (but in a classroom setting) as the system creates an independent path through the material for each student. It assesses students' skills based on how they solve problems rather than on a single end-of-session assessment. The platform captures every decision a student makes while working within lessons. DreamBox provides educators with a robust set of real-time academic progress reports that give teachers and administrators insight into how students are progressing through the intelligent adaptive math curriculum. DreamBox analyses a student's input on *click-by-click basis*, and uses this input to assign the most appropriate follow-up lessons. As a result, there are millions of different "paths" students can take through Dreambox's learning units. This also means that the student is initially presented with only a limited selection of lessons to do. DreamBox will analyse the student's responses to those, and then present him/her new lessons based on the analysis. Another very distinctive feature of DreamBox is its emphasis on conceptual understanding and algebraic thinking[4].

III.   COMPARISON OF EXISTING ONLINE LEARNING SYSTEMS

ASSISTments online platform provides a very fast processing speed as compare to other online learning systems like Wayangoutpost and Dreamboxlearning .While processing speed of the Dreambox learning is very slow, sometimes it takes several minutes for opening a page. Working on the Online learning system we need to have an account so that we can accessed its features but this is not in the case of Wayangoutpost online tutor .In Wayangoutpost we can we can access the learning features of it by logging as a guest.

The working in the ASSISTments is very easy, it has a very user friendly interface with all the features and functionality that a user of it wants everytime working on it.While working on the Wayangoutpost and Dreambox is not so easy as it sometimes create confusion for their users where to go and what to do. Besides an User friendly interface provided by the ASSISTments it also provides an interface where an user learns not in an traditional classroom setting environment where an student feels bored after some time but have an environment of fun while learning and assessing its performance in a motivated sense .The learning interface provided by the Wayangoutpost and Dreambox learning platform gives a sense of a game interface ,so students while working there have a feel of that they are playing a game but the reality is that they are also learning along with assessing their performance. Customization in ASSISTments for individual student based on their performance is not as much as it is in Wayangoutpost and Dreambox .In ASSISTments customization is limited up to the hints used by the student, question sets are pre-built not according to the student skills or requirements .While in the Wayangoutpost, the student's performance is assessed first by offering certain questions and based on the responses of the student ,their cognitive skills are identified after that on the basis of their skills question sets are offered to them .Dreambox Learning identified the learning pace of the student and depending on that offered them new questions for learning.

Table I
Comparison of existing Online Learning Systems

ASSISTments	Wayangoutpost	Dreambox
Processing speed is fast	Processing speed is fast	Processing speed is slow
Need to have an account	Need not to have an account	Need to have an account
User friendly interface	Less user friendly interface	Less user friendly interface
Less Game like interface	Game like interface	Game like interface
Less customized	More customized	More customized

IV. PROPOSED WORK

The Online learning system we will develop is a platform for learning about the courses offered in it covering the one field of analytics that is computer science it has various features that a online learning system must have and overcoming all the shortcomings of existing online learning systems. In the proposed online learning system there is a need that every user must registered before using the features and functionality of the system but some features can be used by the unregistered user.

The proposed online learning system must have

- A well planned and organised program of learning.
- Well-designed, interactive and up-to-date learning resources and assessment materials, which use plain and clear English.
- Opportunities to communicate and interact with teachers and other students using technology (discussion forums, chat rooms)
- Online assessment and feedback: valid reliable, easy to use, responsive and speedy feedback.
- Ease of use: processes and design features which avoid problems such as confusion, excessive processing time for resources and other information, unclear learning structures .
- Online news and announcement
- Online Quiz
- Preprocessing of the moodle data for data mining
- Use of Rapidminer for applying data mining on the moodle data
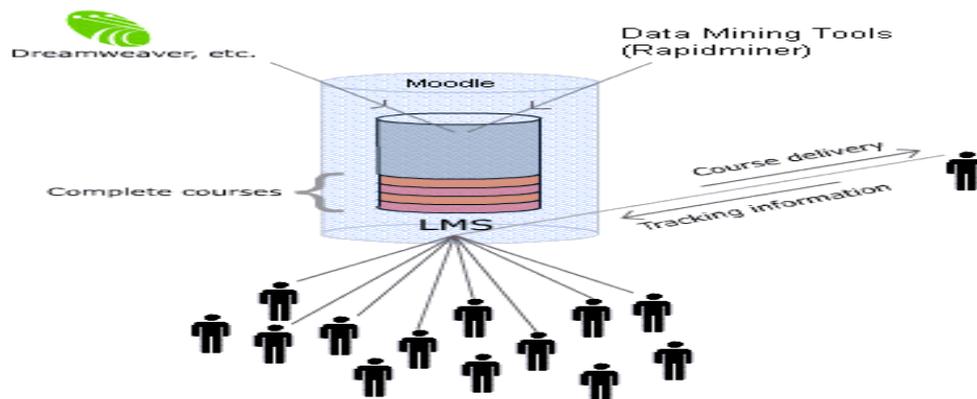- Presenting of the data after applying data mining

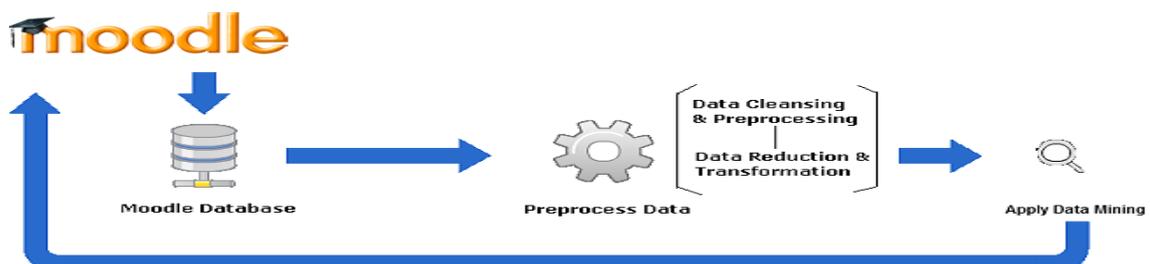Figure.1 Showing the Propsed Project Structure

Figure 2. Showing the processing after the project build and before applying the data mining on the Database
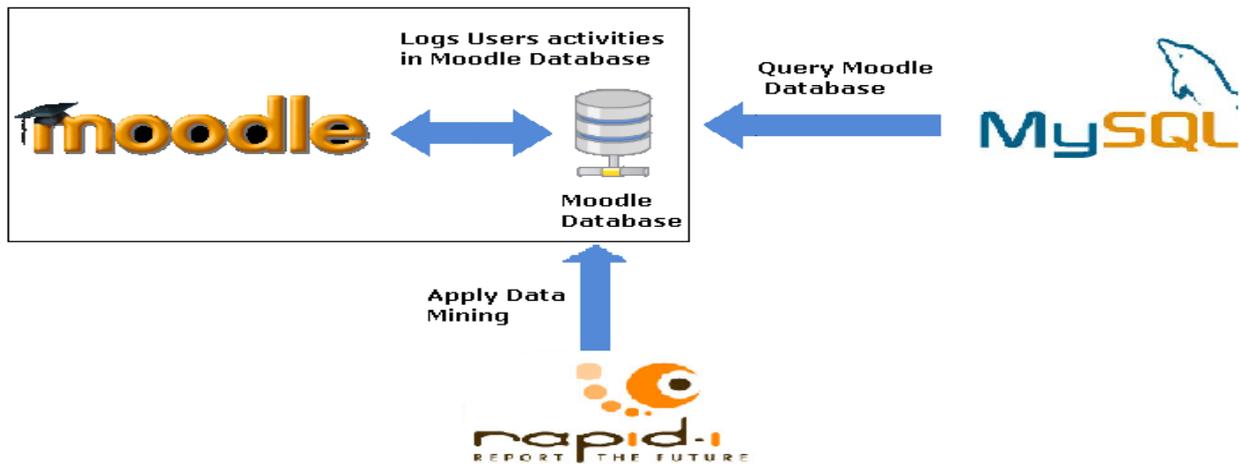
Figure 3.Showing the Data Preprocessing before applying the Data Mining Tool

V. METHODS USED

The LMS system is used by students and the usage and interaction information is stored in the database. In this paper we are going to use the students' usage data of the Moodle system. Preprocess the data. The data is cleaned and transformed into an appropriate format to be mined. In order to preprocess the Moodle data, we can use a database administrator tool or some specific preprocessing tool. Apply data mining. The data mining algorithms are applied to build and execute the model that discovers and summarizes the knowledge of interest for the user (teacher, student, administrator, etc.). In order to do so, we can use a general or a specific data mining tool, and we can use a commercial or free data mining tool .In this paper we use the RapidMiner data mining tool to applied data mining on the moodle data. Interpret, evaluate and deploy 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[5]. Moodle is an open-source learning course management system to help educators create effective online learning communities. Moodle is an alternative to proprietary commercial online learning solutions, and is distributed free under open source licensing. Moodle's modular design makes it easy to create new courses, adding content that will engage learners. Moodle is designed to support a style of learning called social constructionist pedagogy . This style of learning believes that students learn best when they interact with the learning material, construct new material for others, and interact with other students about the material. Moodle does not require the use of this style in the courses but this style is what it best supports. Moodle keeps detailed logs of all activities that students perform [5].

Logging is record keeping that can keep track of what materials students have accessed. Moodle logs every click that students make for navigational purposes and has a modest log viewing system built into it. Log files can be filtered by course, participant, day and activity. The teacher can use these logs to determine who has been active in 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 responses and item analysis of the items themselves. Teachers can easily get full reports of the activities of individual students, or of all students for a specific activity. Activity reports for each student are available and details about each module (last access, number of times read) as well as a detailed story of each student's involvement. Logs can show the activity in the class for different days and times. This can be useful to check to see if everyone has done a certain task, or spent a required amount of time online within certain activities [5].
Moodle does not store logs as text files. Instead, it stores the logs in a relational database. So, data are stored in a single database. MySQL and PostgreSQL are the best supported, but it can also be used with Oracle, Access, Interbase, and others .In this paper we have used MySQL because it is the world's most popular open source database. The Moodle database has about 145 interrelated tables .But we do not need all this information and it is also necessary to convert it into the required format used by the data mining tool and algorithms .For this reason, we have to perform a previous step to preprocess the moodle data. Data preprocessing allows the original data to be transformed into a suitable shape to be used by a particular data mining algorithm or framework.. In order to preprocess the Moodle data, we can use a database management tool such as MySQL Administrator tools , phpMyAdmin , etc. These tools see, edit, delete, modify all the tables and table data, execute SQL queries, create new tables [5].

It is necessary to create a new table in the Moodle database that can summarize the information at the required level (e.g. student). Student and interaction data are spread over several tables. We have created a new performance table which integrates the most important information for our objective. This table has a summary per row about all the obtained by each student in the course in both assignment and quiz. The new table performance created in the Moodle database from other existing tables by executing the several queries to the database in order to obtain the required student's details. After creating the table, data mining tool is used to apply mining over the moodle data .In this paper, we uses the Rapid Miner .

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. A large numbers of operators (more than four hundred) defined in RapidMiner along with its plugins cover nearly all key aspects of Data mining handling data transparently and without the need to know the different data formats or different data views. 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. RapidMiner follows a multi-layered data view concept which enables it to store different views on the same data table and therefore facilitates cascading multiple views in layers through a central data table. RapidMiner data core is typically similar to a standard database management system. RapidMiner has a flexible interactive design which lets user to additional meta data on the available data sets to enable automated search and optimized preprocessing which are both needed for an effective data mining processes.

Table II
Some important Moodle tables for doing data mining

| Name	Description
mdl_user	Information about all the user
mdl_user_students	Information about all students.
mdl_log	Logs every user's action.
mdl_assignement	Information about each assignment. .
mdl_assignment_submissions	Information about assignments submitted.
mdl_chats	Information about all chatrooms.
mdl_glossary	Information about all glossaries.
mdl_forum	Information about all forums.
mdl_forum_posts	Stores all posts to the forums.
mdl_quiz	Information about all quizzes.
mdl_quiz_attempts	Stores various attempts at a quiz
mdl_quiz_grades	Stores the final quiz grade

VI. RESULTS
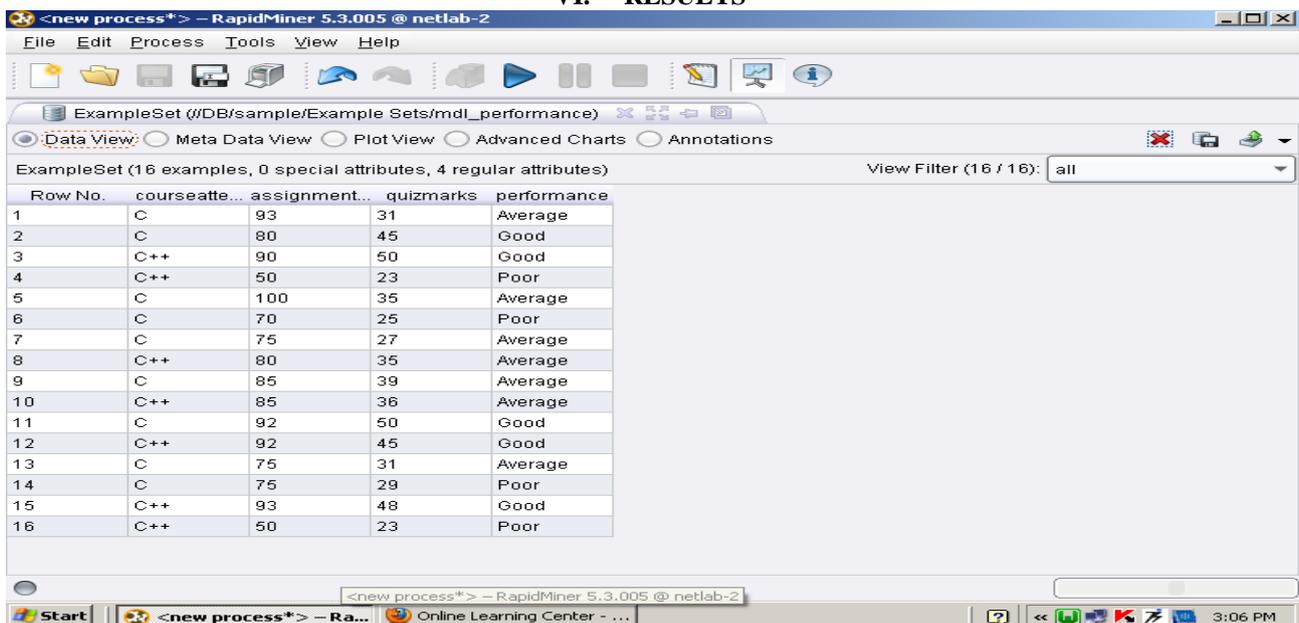
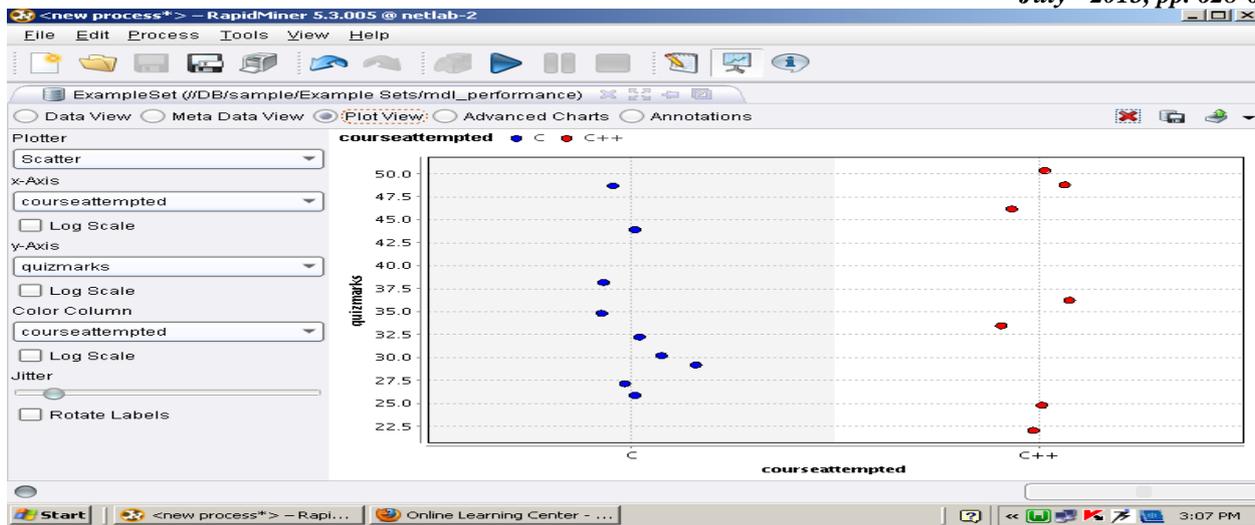Figure 4. Showing the look of the new table performance and it's contents.

Figure 5. Showing the figure below shows the table performance from plot view in the RapidMiner tool.

In this paper we uses the Classification technique of data mining and apply decision tree algorithm for evaluating the student's performance.The below figure shows the application of decision tree algorithm for predicting student performance.The performance table is used throughout the process for predicting student's performance.
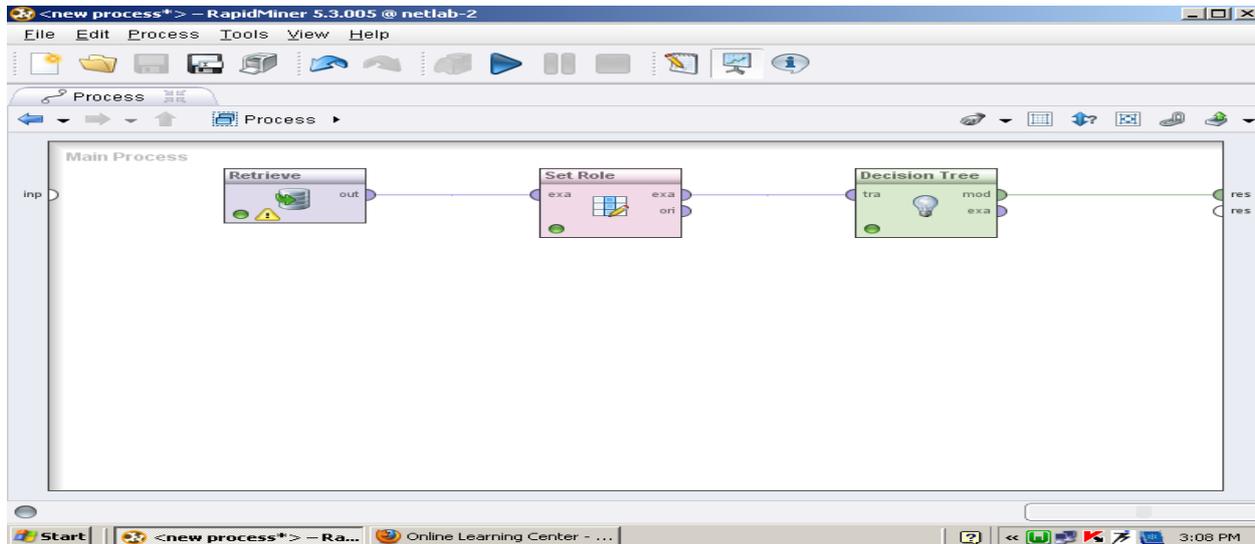

Figure 6. Showing the application of decision tree algorithm for predicting student performance.
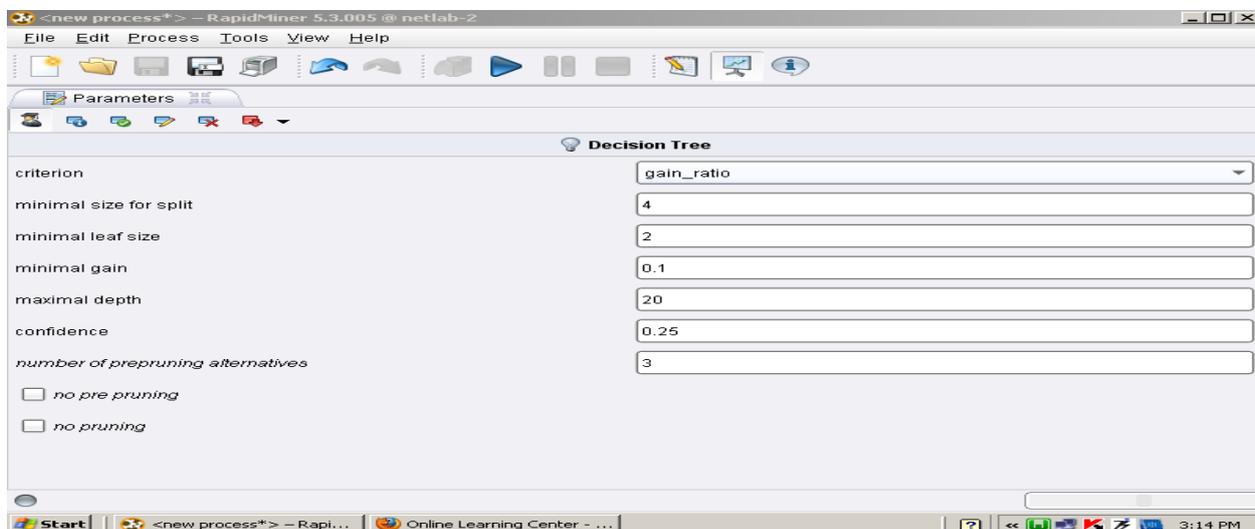
Figure 7.Showing the parameter settings of the decision tree operator ,we set the criterion of the decision tree operator to gain ratio.

The figure 8 shows the result of the applying decision tree algorithm over the performance table in the decision tree form.
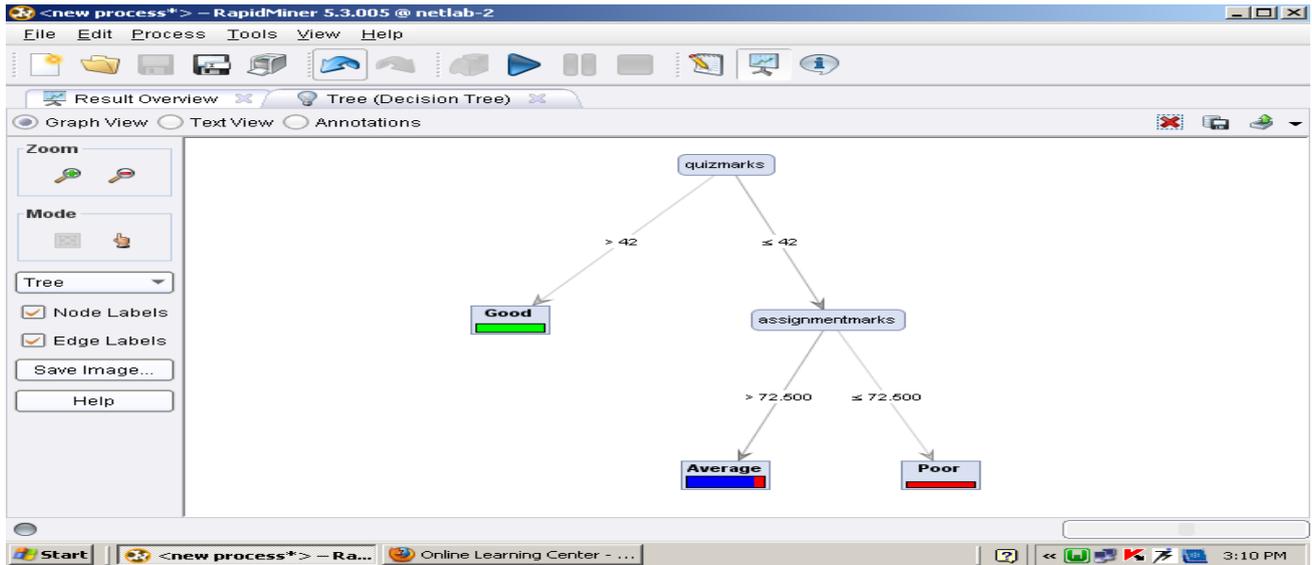

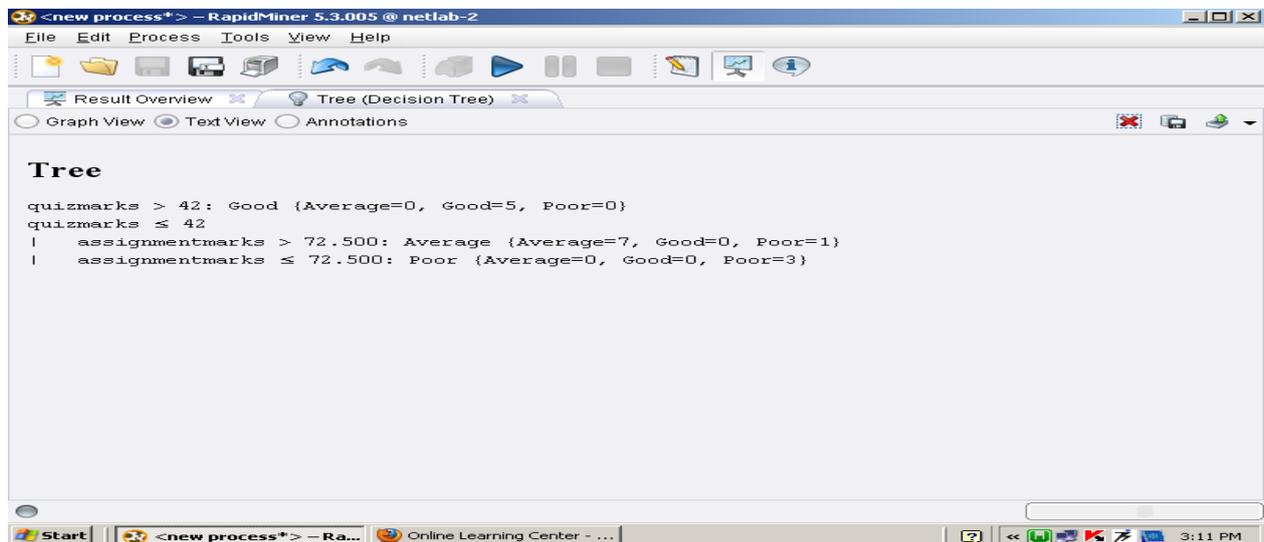Figure 8. Showing the result of data mining application.

Figure 9 . Showing the above result is shown in the text view.

In the below figure,we shows the change parameter settings of the decisiocn tree operator now the criterion is set to Information_gain for getting more expanded form of the data miningresult.
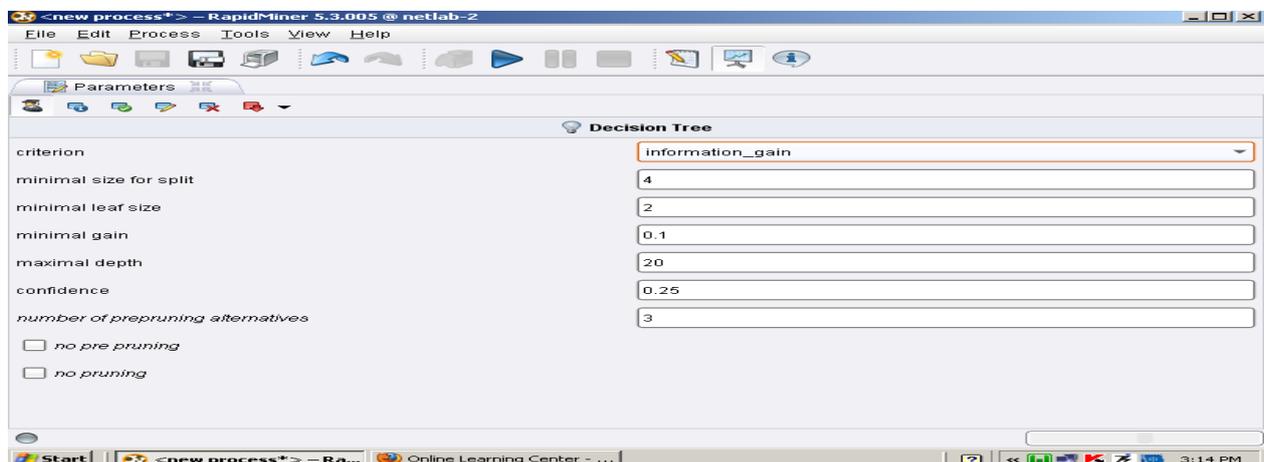

Figure 10. Showing the parameter settings

The below figure shows the result of the appling decision tree algorithm over the performance table in the decision tree form.
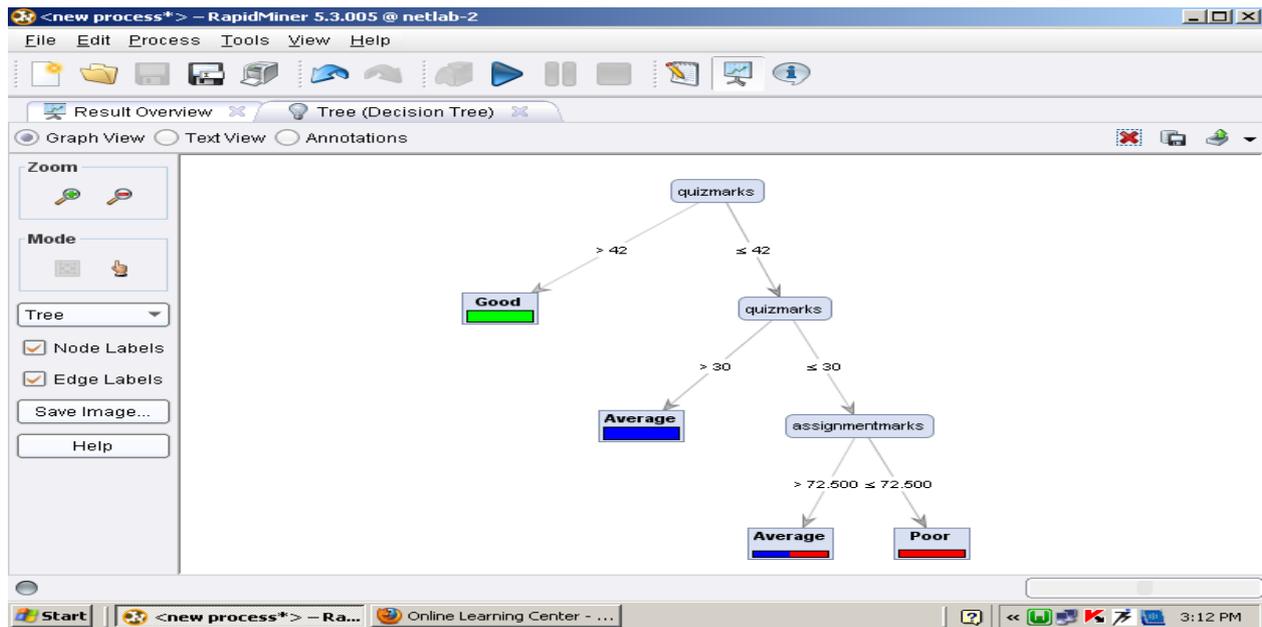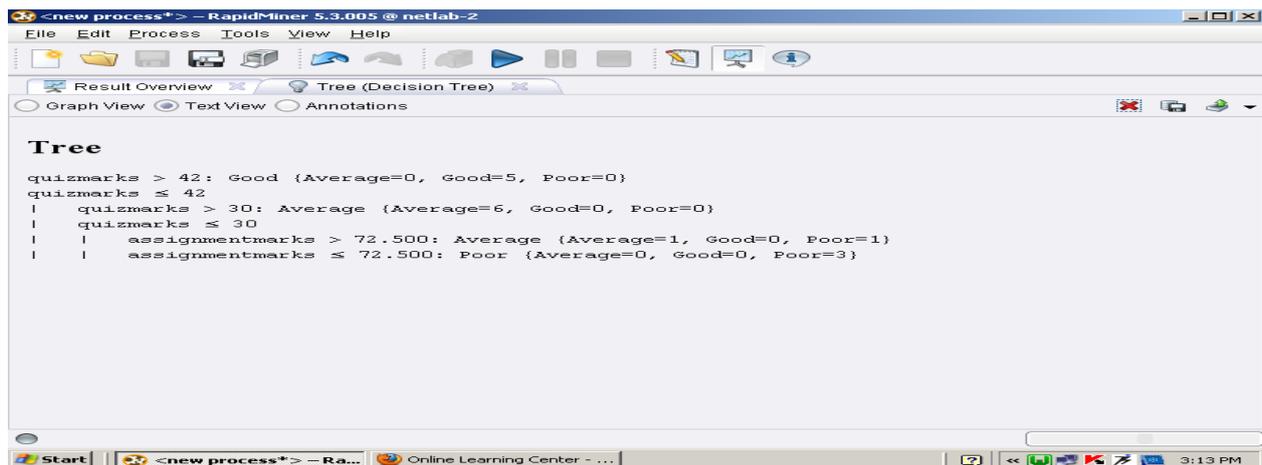

Figure 11.Showing the result of applying data mining.

Figure 12. Showing the above result in the text view

VII.   CONCLUSION

In this study we make a use of data mining process in a moodle database using data mining tool to predict students performance .We hope that the information generated after the implementation of data mining technique may be helpful for a instructor as well as for students.

REFERENCES
[1].   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.
[2].   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.
[3].   http://www.assistments.org/
[4].   Arne Duncan and Karen Cator,"Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics", U.S Department of Education, 2012.
[5].   Cristobal Romero, Sebastian Ventura and Enrique Garcia,"Data Mining in course management system : Moodle case study and tutorial".