# DATA MINING AND E-LEARNING PROCESS

Prakash Chandra Behera[1]        Chinmaya Dash[2] Somanjoli Mohapatra[3]

[1]Assistant Professor, Department of Sciences, St. Claret College, Bangaloe-560013
[2]Assistant Professor, Department of Sciences, St. Claret College, Bangaloe-560013
[3]Assistant Professor, Department of Sciences, St. Claret College, Bangaloe-560013

## Abstract:

Data mining is a great innovative technology which helps corporations to focus on the most important information in the data of their stockrooms. Data mining is used in various statistical, machine learning and in graphical methods. Data mining separates the knowledge in to a form which is very useful for many real world applications. Educational Data Mining (EDM) is the process of converting raw data from Educational systems to useful information that can be used by educational software Developers, students, teachers, parents, and other educational researchers. Currently there is an increasing interest in data mining and educational systems, making educational data mining as a new growing research community. This work provides the application of data mining to traditional educational systems, particular web-based courses, well-known learning content management systems, and adaptive and intelligent web-based educational systems. Each of these systems has different data source and objectives for knowledge discovering. After preprocessing the available data in each case, data mining techniques can be applied: statistics and visualization; clustering, classification and outlier detection; association rule mining and pattern mining; and text mining. The success of the plentiful work needs much more specialized work in order for educational data mining to become a mature area.

*Keywords: Data mining and E-learning, Educational Data Mining,Process of Data Mining in E-Learning*

## INTRODUCTION

E-learning (also referred to as web-based education and eteaching), a new context for education where large amounts of information describing the continuum of the teaching-learning interactions are endlessly generated and ubiquitously available. This could be seen as a blessing: plenty of information readily available just a click away. But it could equally be seen as an exponentially growing nightmare, in which unstructured information chokes the educational system without providing any articulate knowledge to its actors, Data Mining was born to tackle problems like this. As a field of research, it is almost contemporary to e-learning. It is, though, rather difficult to define. Not because of its intrinsic complexity, but because it has most of its roots in the ever-shifting world of business. At its most detailed, it can be understood not just as a collection of data analysis methods, but as a data analysis process that encompasses anything from data understanding, preprocessing and modeling to process evaluation and Implementation. It is nevertheless usual to pay preferential attention to the Data Mining methods themselves. These commonly bridge the fields of traditional statistics, pattern recognition and machine

learning to provide analytical solutions to problems in areas as diverse as biomedicine, engineering, and business, to name just a few. An aspect that perhaps makes Data Mining unique is that it pays special attention to the compatibility of the modeling techniques with new Information Technologies (IT) and database technologies, usually focusing on large, heterogeneous and complex databases. Learning databases often fit this description.

## Data Mining

Data mining is an interactive process within which progress is defined by discovery through either automatic or manual methods. Businesses can learn from their transaction data about the behavior of the customers and thus can improve their business by exploiting this knowledge Web usage information can be analyzed and exploited to optimize information access. Thus data mining generates novel, unsuspected, unbiased, interpretations of data. The main idea of data mining falls under 2 categories .

a. Predictive data mining: creates the model of the system from the given data.

 b. Descriptive data mining: This generates significant data sets from the existing data.

The aim of these above ideas is achieved by the following data mining techniques.

1. Characterization- Characterization is used to generalize, summarize and possibly different data characteristics.
2. Classification-Data classification is a process in which the given data is classified in to different classes.
3. Regression-This process is similar to classification. the major difference is that the object to be predicted is continuous rather than discrete.
4. Association-It discovers the association between various data bases and the association between the attributes of single database.
5. Clustering-Clustering involves grouping of data into several new classes such that it describes the data. It breaks large data set into smaller groups to make the designing and implementation process to be simple. The task of Clustering is to maximize the similarity between the objects of classes and to reduce the similarity between the classes.
6. Change Detection-This method identifies the significant changes in the data from the previously measured values.
7. Deviation Detection-Deviation detection focuses on the major deviations between the actual values of the objects and its expected values. This method finds out the deviation according to the time as well the deviation among different subsets of data.
8. Link Analysis-It traces the connections between the objects to develop models based on the patterns in the relationships by applying graph theory techniques.
9. Sequential Pattern Mining-This method involves the discovery of the frequently occurring patterns in the data.

## E-learning

E-learning (also referred to as web-based education and e-teaching), a new context for education where large amounts of information describing the continuum of the teaching-learning interactions are endlessly generated and ubiquitously available. This could be seen as a blessing: plenty of information readily available just a click away. But it could equally be seen as an exponentially growing nightmare, in which unstructured information chokes the educational system without providing any articulate knowledge to its actors, Data Mining was born to tackle problems like this. As a field of research, it is almost contemporary to e-learning. It is, though,

rather difficult to define. Not because of its intrinsic complexity, but because it has most of its roots in the ever-shifting world of business. At its most detailed, it can be understood not just as a collection of data analysis methods, but as a data analysis process that encompasses anything from data understanding, preprocessing and modeling to process evaluation and Implementation.

## Data mining and E-learning

Data mining and E-learning Aims to provide an up-todate snapshot of the current State of research and applications of Data Mining methods in e-learning. The Cross-fertilization of both areas is still in its infancy, and even academic References are scarce on the ground, although some leading education-related Publications are already beginning to pay attention to this new field. In order to Offer a reasonable organization of the available bibliographic information According to different criteria, firstly, and from the Data Mining practitioner Point of view, references are organized according to the type of modeling Techniques used, which include: Neural Networks, Genetic Algorithms, Clustering and Visualization Methods, Fuzzy Logic, Intelligent agents, and Inductive Reasoning, amongst others. From the same point of view, the Information is organized according to the type of Data Mining problem dealt with: clustering, classification, prediction, etc. Finally, from the standpoint of the e-learning practitioner, we provide taxonomy of e-learning problems to Which Data Mining techniques have been applied, including, for instance: Students' classification based on their learning performance; detection of Irregular learning behaviors; e-learning system navigation and interaction Optimization; clustering according to similar e-learning system usage; and systems' adaptability to students' requirements and capacities.

Educational Data Mining (called EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. the areas of EDM application are

- Analysis and visualization of data
- Providing feedback for supporting instructors
- Recommendations for students
- Predicting student performance
- Student modeling
- Detecting undesirable student behaviors
- Grouping students
- Social network analysis
- Developing concept maps
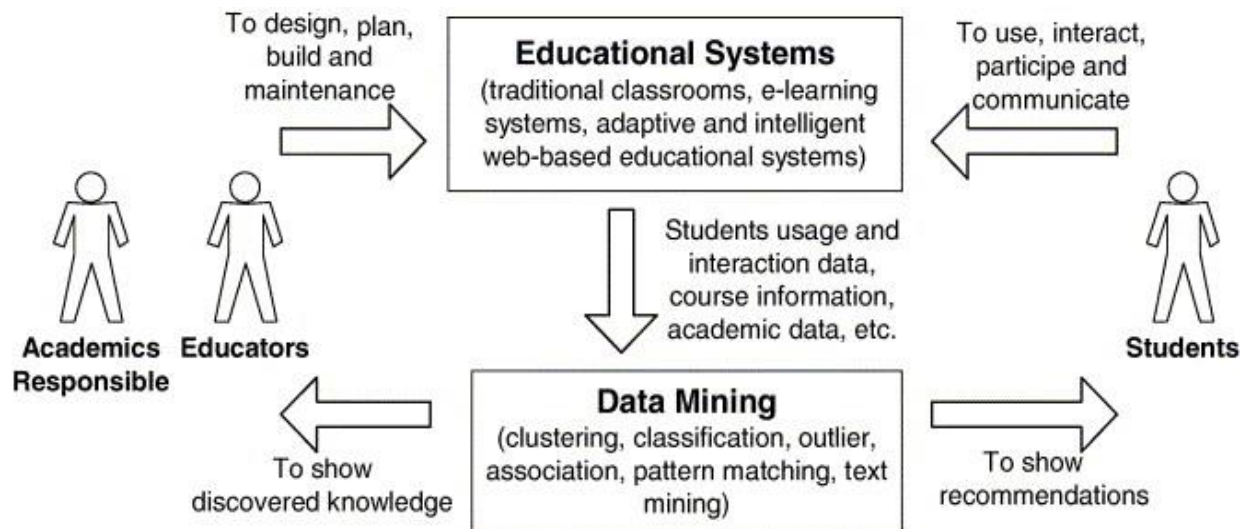- Constructing courseware
- Planning and scheduling

Fig 1: Structure of Data Mining and e-learning

## Data mining functionalities for e-learning domain:

In the e-learning domain, we are interested in managing mainly two groups of users: the learners as well as the learning providers, whether private training companies, governmental organizations and local authorities providing training for their employees or universities who aim to publish their courses and make them accessible online via the Internet. As for learners, databases should store all personal details including name, age, gender, address, postcode, and educational-relevant details such as qualifications. Moreover having information like work experience, career objectives, income range, previous courses taken and courses of interest would be of great value to be able to predict future behavior of different classes of employed professional people. Also other information such as personal interests and hobbies would be very valuable for data mining tool in order to discover hidden patterns by building intelligent models based on the huge amount of data Web usage mining can be applied to e-learning domain as the site records information Recording learner profiles, web access information, academic details of students and evaluation results. Web usage mining can track learning activities and identifies web access patterns and user behaviors. Web usage mining has lot of contributions in e-learning domain such as

- (i)    Dynamic personalization like providing real time recommendations for e-learners
- (ii)   Commonly referenced web pages are cached in proxy servers.
- (iii)  Structuring or organizing the site structure according to learner's interest.
- (iv)   Creating access shortcuts for interested pages to enhance user friendliness.
- (v)    Updating course content of web site according to previous usage information.
- (vi)   Identifying groups of learners of similar interest and sending personalized course materials to interested groups.

## Processes of Data Mining in E-Learning

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.The application of data mining in educational systems has specific requirements, mainly the need to take into account learners' specific behavior, including pedagogical aspects. The application of data mining in E-learning systems can be described as an iterative cycle where

data mining applications contribute in enhancing learning, and also using mined knowledge for decision making.The e-learning data mining process consists of the same four steps in the general data mining process as follows.
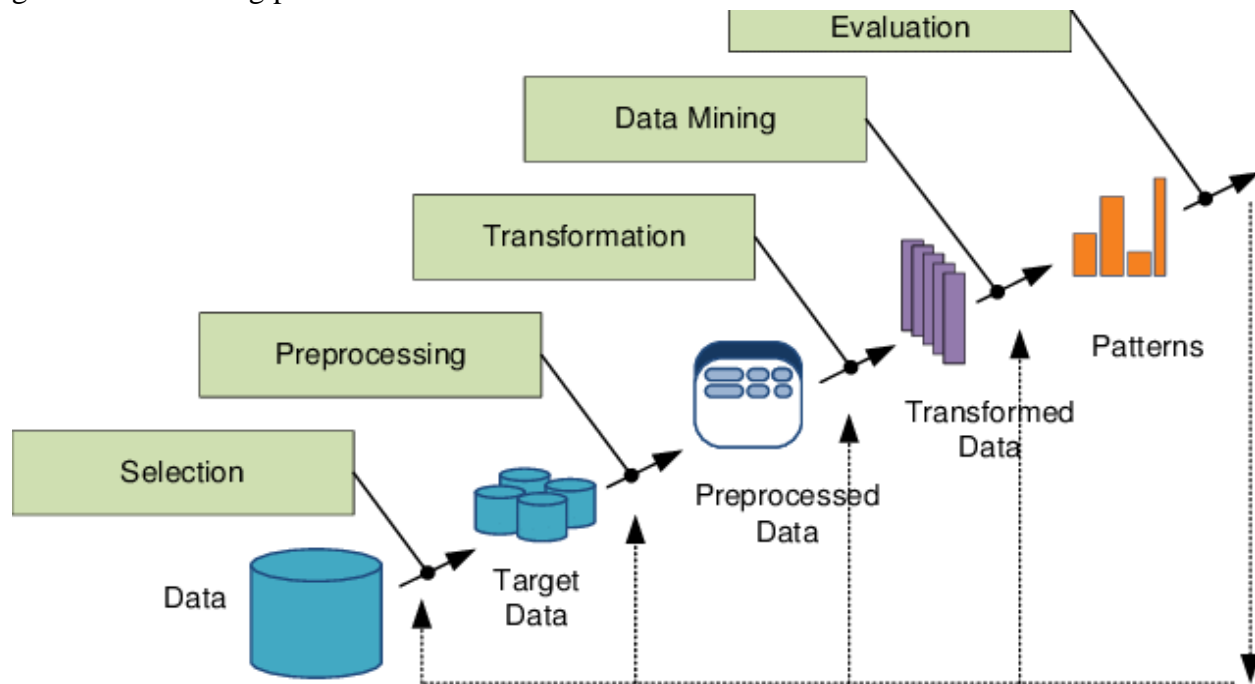


Fig 2: Architecture of data Mining

1. Collect data. The LMS system is used by students and the usage and interaction information is stored in the database. Interaction information is stored in the database of the LMS.

2. Preprocess the data. The data is cleaned and transformed into an appropriate format to be mined. The data is transformed into an appropriate format.

3. 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. The data mining algorithms are applied to create and execute the model that discovers the knowledge and patterns of interest. In order to achieve this goal a data mining tool can be used.

4. Interpret, evaluate and deploy the results. The results or model obtained are interpreted and used by the teacher for further actions. The educator can use the information discovered to make decisions about e-leaning system and process.

## The application of data mining in e-learning systems

Data mining can be used to resolve classification problems in e-learning. Only a few data mining techniques can be applied to elearning to resolve classification problem. The techniques are: fuzzy logic methods; artificial neural networks and evolutionary computation, graphs and trees; association rules; multi-agent and systems. Furthermore, application of data mining to resolve clustering problems in e-learning; includes artificial neural network and clustering.

Data mining application in e-learning is an iterative cycle 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.

## Conclusion

Educational data mining's an upcoming field related to several well-established areas of research including e-learning, adaptive hypermedia, intelligent tutoring systems, web mining, data mining, etc. The application of data mining in educational systems has specific requirements not present in other domains, mainly the need to take into account pedagogical aspects of the learner and the system.Mining tools more easy to use by educators or not expert users in data mining. Data mining tools are normally designed more for power and flexibility than for simplicity. Most of the current data mining tools are too complex to use for educators and their features go well beyond the scope of what a educator may want to do.Standardization of methods and data - current tools for mining data from a specific course may be useful only to its developers. There are no general tools or re-using tools or techniques that can be applied to any educational system. So, a standardization of data, and the preprocessing, discovering and post processing tasks is needed.Integration with the e-learning system - The data mining tool has to be integrated into the e-learning environment as another author tool. All data mining tasks (preprocessing, data mining and post processing) have to be carried out into a single application. Feedback and results obtained with data mining can be directly applied to the e-learning environment.

Specific data mining techniques- More effective mining tools that integrate educational domain knowledge into data mining techniques. Education-specific mining techniques can help much better to improve the instructional design and pedagogical decisions. Traditional mining algorithms need to be tuned to take into account the educational context.

## IX. References

[1] Anduela Lile,Mediterranean Journal of Social Sciences," Analyzing E-Learning Systems Using Educational Data Mining Techniques", (2011).

[2] Félix Castro1, 2, Alfredo Vellido1, Àngela Nebot1, and Francisco Mugica3, "Applying Data Mining Techniques to eLearning Problems",(2007)

[3] C. Romero., S. Ventura," Data Mining in E-Learning", University of Cordoba, Spain, University of Cordoba, Spain,(2006).

[4] Margo Hanna, "Data mining in the e-learning domain", Campus-Wide Information Systems, (2004)

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

[6] "EducationalDataMining.org", (2010)

[7] R. Baker,"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: Elsevier, (2010)

[8]A.Anitha, Dr.N.Krishnan" A Dynamic Web Mining Framework for E-Learning Recommendations using Rough Sets and Association Rule Mining',(2011).

[9] Jean Baptiste Batware" REAL-TIME DATA MINING FOR E-LEARNING",( 2007 )

[10] SUN, P., TSAI, J., FINGER, G., CHEN, Y. & YEH, D. What drives a successful E-learning? An empirical investigation of the critical factors influencing learner satisfaction (2007)