

Applications and Techniques of Educational Data Mining in the Field of Higher Education (A review on Data Mining and Educational Data Mining in Education Sector)

Charushila D. Patil, Vishkha N.Pawar

Head of Department, Department of Computer Engineering, Guru Gobind Singh Polytechnic, Nashik, India

Lecturer, Department of Computer Engineering, Guru Gobind Singh Polytechnic, Nashik, India

ABSTRACT: The biggest challenge for today's education sector is effectively managing all the data of students staff, faculties as well as various facility which they are using in their regular academics by effective and efficient way. This raw data is very important, but managing all this data is very tedious job if we have to handle it manually. Data mining is one of the effective techniques of extracting meaning data a from large database. a Educational data mining is one of the sub branch of data mining where all the data related to result, performance, behavioral analysis of students is extracted. This paper gives a brief overview of Educational Data mining, its objectives and various techniques for higher education.

KEYWORDS: Educational Data Mining (EDM); Classification, K-Means Alogorithm

I. INTRODUCTION

Data mining is a vast technique which searches retrieves and analyze data from a data repository say as warehouse which is required for business purpose.

It is the process of studying data from different perceptions and briefing it into useful information. The information obtained from data mining can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of investigative tools used for searching and testing data. It allows users to analyze data from many different scopes or angles, classify it, and summarize the relationships identified among the retrieved data . Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases [1]

Now a day, the educational institutes have to require financial assistance for their research and other operational activities as the government funding has been limited to aided institutes.

Employing a decision support system is a positive way to use data to manage, operate, and evaluate educational institute in a better way. Depending on the quality and availability of the underlying data, such a system could address a wide range of problems by cleansing data from any combination of education records maintenance system. Mining data from data warehouse can be a ready and effective system for the decision makers.

data warehouse is a subject oriented, integrated, time-variant, and non-volatile collection of data. This data helps experts to take knowledgeable decisions in an organization.[5]

Features of Data Warehouses:

- It is a database, which is kept separate from the organization's operational database.
- No regular updating done in a data warehouse.
- It possesses consolidated ancient data, through which organization can analyze its business.
- It helps executives to organize, understand, and use their data to take strategic decisions.
- It helps in the integration of diversity of application systems.[5]

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 6, Special Issue 1, January 2017

International Conference on Recent Trends in Engineering and Science (ICRTES 2017)

20th-21st January 2017

Organized by

Research Development Cell, Government College of Engineering, Jalagon (M. S), India

II. MOTIVATION

The need of building data warehouse for the educational institute is generated due to following two main reasons:

- 1) internal sources:
- 2) External Sources

Internal Sources includes incapability of current operational systems to provide required information for parameter driven analysis external sources include competitive factors.

Generally following problems are faced by all educational Institutes:

- The data is stored in different sources in distributed locations.
- Users find difficulty in locating the reports needed by them.
- The user interface for the current operational system is not satisfactory and is confusing and hard to use for decision makers.
- When the consolidated report from two or more different subject area is required, it is almost impossible.
- There is no easy way to get assistance.

Thus by using a decision support system is a positive way to use data to manage, operate, and evaluate educational institute in a better way. Depending on the quality and availability of the underlying data, such a system could address a wide range of problems by cleansing data from any combination of education records maintenance system.

III. GOALS OF EDM

Following are the four goals of EDM.

1. Predicting students' future learning behavior- By using student modeling, this goal can be achieved by creating student models that incorporate the learner's characteristics, including detailed information such as their knowledge, behaviors and motivation to learn. The user experience of the learner and their overall satisfaction with learning are also measured.
2. Discovering or improving domain models – By using different method of EDM, discovery of new and improvements to existing models is possible. Examples include illustrating the educational content to engage learners and determining optimal instructional sequences to support the student's learning style.
3. Studying the effects of educational support that can be achieved through learning systems.
4. Advancing scientific knowledge about learning and learners by building and incorporating student models, the field of EDM research and the technology and software use

IV. EDUCATION DATA MINING

Educational data mining (EDM) is a branch of data Mining which develops, find out as well applies computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of data within which they exist. EDM has emerged as a research area in recent years aimed at analyzing the unique kinds of data that arise in educational settings to resolve educational

A. USERS AND STAKEHOLDERS OF EDM

EDM is constituted with four main users and stakeholders. These include:

Learners – these are interested in understanding student needs, their views various methods to improve the learner's experience and performance. For example, learners can also get advantage from the found knowledge by using the EDM tools to suggest activities and resources that they can use based on their interactions with various online learning techniques and software tool and perceptions from past or similar learners. For younger learners, educational data mining can also inform parents about their child's learning progress It is also necessary to effectively group learners in

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 6, Special Issue 1, January 2017

International Conference on Recent Trends in Engineering and Science (ICRTES 2017)

20th-21st January 2017

Organized by

Research Development Cell, Government College of Engineering, Jalagon (M. S), India

an online environment. The challenge is to learn these groups based on the complex data as well as develop actionable models to interpret these groups

Educators – They tries to get knowledge about the learning process and the methods they can use to improve their teaching methods they can use the applications of EDM which concludes method of organizing and structuring the curriculum, the best methods to deliver course information and the tools to use to engage their learners for optimal learning outcomes.

Researchers – these people mainly focus on the development and the evaluation of data mining techniques for effectiveness

Administrators – these are responsible for allocating the resources for implementation in institutions As student success is very crucial factor, institute is more concern about them and Administering of EDM applications are becoming more common in educational settings. Faculty and advisors are becoming more positive in recognizing and addressing at-risk(threatened) students.

In Academics, a student’s performance is determined by the term work, attendance and end semester examination.

To carry out term work, teacher needs individual student performance measurement. these types of assessment is done computerized by technics of EDM.

B. TECHNIQUES OF EDUCATIONAL DATA MINING:

Techniques	Description
Classification	Classifying and summarizing students, determining their learning styles and preferences
Predictive Modeling	Making models that can predict when and how a student will pass a course
Clustering	Grouping Similar students (based on behavior, performance)or grouping similar courses, assignments etc together, discovering collaborative learning patterns
Bi-Clustering	Searching nature of question, and it’s difficulty level
Frequent Pattern Mining	Finding (elective) courses often taken together

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 6, Special Issue 1, January 2017

International Conference on Recent Trends in Engineering and Science (ICRTES 2017)

20th-21st January 2017

Organized by

Research Development Cell, Government College of Engineering, Jalagon (M. S), India

V.APPLICATIONS OF DATA MINING IN HIGHER EDUCATION

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

Some applications are described as follows:

A. ANALYSIS AND VISUALIZATION OF DATA:

It is used to focus meaningful information and support decision making.eg. course administrators and educators will analyze the usage information and students' activities during course to get a brief idea of a student's learning. Visualization information and statics are the two main methods that have been used for this task. the Detailed mathematical analysis of educational data can give us information like where students enter and exit, the most important pages students browse, how many number of downloads of e-learning resources, quantity of pages browsed and total amount of time for browsing of these different pages. It also provides information about reports on monthly and weekly usage of tendencies, usage summaries, material studied by students, the series in which they study topics, patterns of studying activity, timing and sequencing of activities.

B. PREDICTING STUDENT PERFORMANCE

Here, we predict the unknown value of a variable that defines the student. In educational sector, the mostly predicted values are student's performance, are the marks they achieved, knowledge they gained or thus overall score.

Classification technique is used to combine individual items based upon quantitative traits or based on training set of previously labeled items.

Various techniques for predicting student's performance are: decision trees, neural networks, rule based systems, Bayesian networks etc.

C. ENROLMENT MANAGEMENT:

It regularly used in higher education to explain well-planned strategies and ways to shape the enrolment of college to meet planned goals. Such practices often include retention programs, marketing, financial assistance and admission policies.

D. Grouping Students:

In this case groups of students are created according to their customized features, personal characteristics, etc. These clusters of Students can be used by the teacher for building a personalized learning system which can promote effective group learning.

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 6, Special Issue 1, January 2017

International Conference on Recent Trends in Engineering and Science (ICRTES 2017)

20th-21st January 2017

Organized by

Research Development Cell, Government College of Engineering, Jalagon (M. S), India

The Data Mining techniques used in this task are classification and clustering. Various clustering algorithms used to group students are as follows:

- hierarchical agglomerative clustering,
- K means
- Model based clustering.

D. PREDICTING STUDENTS

Summarizing Data mining can help management to identify the demographic, geographic and psychographic characteristics of students based on information provided by the students at the time of admission. Techniques of DM is used here is neural network

E PLANNING AND SCHEDULING:

Planning and scheduling is used to enhance the traditional educational process by planning future courses, course scheduling, planning resource allocation which helps in the admission and counseling processes, developing curriculum, etc. Following Data mining techniques used:

- classification
- categorization
- estimation
- Visualization.
- Decision trees
- link analysis and decision forests

in course planning to analyze students' course Preferences

F. STUDENT MODELING:

It contains what a learner knows, what the user experience is like, what a learner's behavior and Motivation is, and how satisfied users are with online learning. EDM can be applied in modeling user knowledge user behavior and user experience.

VII. ALGORITHMS USED FOR EDM

A. Clustering Algorithm:

Clustering is a grouping of data division of data into groups of similar objects

It is very crucial in data mining applications such as information retrieval and text mining, scientific data exploration, web analysis, spatial database applications, medical diagnostics, marketing etc.

Data Clustering is unsupervised and statistical data analysis technique. It is used to divide similar type of data into homogeneous groups

Following algorithms are used in education mining in Clustering.

K-Mean Clustering Algorithm

The K-means is one of the best clustering algorithms in data mining.

K-Means is a non-hierarchical clustering

method that seeks to partition the data into the form of one or more clusters. This method partitions the data into clusters so that the data having the same characteristics are grouped into one cluster and the data that have different

VII. CONCLUSION

Thus this paper gives a brief overview on Educational Data mining various techniques and applications used in Educational sector for improving quality of higher Education.

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 6, Special Issue 1, January 2017

International Conference on Recent Trends in Engineering and Science (ICRTES 2017)

20th-21st January 2017

Organized by

Research Development Cell, Government College of Engineering, Jalagon (M. S), India

REFERANCES

- [1]<http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>
- [2] Manjunath T. N., Ravindra S. Hegadi, Umesh I. M., and Ravikumar G. K., "Realistic Analysis of Data Warehousing and Data Mining Application in Education Domain "
- [3] Cristobal Romero* and Sebastian Ventura, "Data mining in education "
- [4] B. Namratha and Anurag Group of Institutions, "Educational Data Mining –Applications and Techniques "
- [5]https://www.tutorialspoint.com/dwh/dwh_overview.htm