ORIGINAL RESEARCH PAPER

Volume-8 | Issue-8 | August - 2019 | PRINT ISSN No. 2277 - 8179

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

A STUDY ON EDUCATIONAL DATA MINING: APPLICATIONS AND TOOLS

Computer Science

Mehta Smruti Hemantkumar Research Scholar, Pacific Academy of higher Education and Research University, Udaipur

Dr. Nimesh I. Modi* Department of Computer Science, Hemchandracharya North Gujarat University, Patan
*Corresponding Author

ABSTRACT

Data is most important part of any organization which is used to produce useful information. To take decision from large amount of data is very tedious task. For that various data mining tools are available in market which is useful in finding pattern recognition and support decision making process using these tools. There are various applications of data mining in various fields like education, scientific and engineering, healthcare, business and many more. This paper mainly describes the applications of data mining in the field of education and it also focuses on education data mining tools.

KEYWORDS

Educational data mining, Educational data mining applications, Educational data mining tools

1. Educational Data Mining

EDM is a broader term which includes areas that directly impact students. Other area includes analysis of educational processes including admissions, alumni relations and course selections. The institutional effectiveness and EDM is concerned in developing various techniques or methods, applications and tools for mining the different types of data that come from educational area. Those methods are useful to better understand the students. Predicting student's performance and mining enrollment data is the most important area of EDM. This paper is organized as follows:

1. Applications of data mining in education field
2. Educational data mining tools

2. EDM Applications

Several Applications and task have been determined to develop the academic and administrative sections of institutions through data mining techniques [1]. The primary applications of EDM are provided by Cristobal Romero and Sebastian Ventura. The areas of EDM application are [2]:

1) Analysis and visualization of data
2) Providing feedback for supporting instructors
3) Recommendations for students
4) Predicting student performance
5) Student modeling
6) Detecting undesirable student behaviors
7) Grouping students
8) Social network analysis
9) Developing concept maps
10) Constructing courseware
11) Planning and scheduling

There are many application areas of data mining but the main fretful area is about data mining applications in educational systems. EDM is most promising area, that is concerned in developing methods for exploring the unique types of data that come from educational settings and those methods are used to better understand students and the settings in which they learn. The main applications of EDM are listed as follows:

1) Analysis and visualization of data:

It is used to highlight useful information and support decision making. In the educational area, for example, it can be helpful to both educators and course administrators in analyzing the usage information and students' activities to get a broad idea of a student's learning. There are two main techniques statistics and visualization which are used for this task. Statistics is concerned with mathematical science which contains collection, analysis, interpretation or explanation and presentation of data. Statistical analysis can provide information like where students enter and exit, the most popular pages students browse, number of downloads of e-learning resources, number of different pages browsed and total time for browsing different pages.

It also provides reports on monthly or weekly user trends, usage summaries, and number of materials students will study and the series to study those topics, patterns of studying activity, timing and sequencing of events.

2) Predicting student performance:

In predicting student performance, we forecast the unknown value of a variable that describes the student. In education area, generally predicted values are student's performance, their marks, knowledge or score. These values can be constant or distinct. Classification is used to cluster individual items based upon quantitative characteristics or training set of previously labeled items. The most popular application of data mining in education is predicting student performance. Different techniques and models are applied in predicting student's performance like decision trees, neural networks, rule based systems, Bayesian networks etc. This analysis is helpful in predicting student's performance i.e. to predict student's success in a course and to predict student's final grade on the basis of features extracted from logged data.

Student's marks are predicted using regression techniques such as linear regression to predict student's academic performance, stepwise linear regression to predict time spent by a student on a learning page, multiple linear regression in identifying variables that can be helpful to predict success in colleges courses and to predict exam results in distance education courses.

3) Grouping students:

In this case groups of students are arranged according to their customized features, personal characteristics, etc. these groups of students can be used by the developer to build a personalized learning system which can promote effective learning. Classification and clustering techniques are used in this task. Different clustering algorithms that are used in grouping students are hierarchical agglomerative clustering, K-means and model-based clustering.

A clustering algorithm is based on large generalized sequences which help in finding groups of students with similar learning characteristics like hierarchical clustering algorithm. This algorithm is used in intelligent e-learning systems to group students according to their individual learning style preferences.

4) Enrollment management:

The phrase enrollment management is generally used in higher education to make clear well-planned strategies and plans to figure out the enrollment of an organization to meet their standard objectives. It is a managerial concept and also a systematic set of activities designed to allow educational institutions to exert (use) more influence (authority) over student's enrollments. This includes marketing, retention program, admission policies and financial aid awarding.

3. Educational Data Mining Tools:

EDM tools can be categorized in two ways. First category consists of open source tools and second category consists of commercial education data mining tools [3].

Table 1 : Open source educational data mining tools

Name of Tool and developer	Function/Features	Techniques	Environment
See5 and C5.0	Provides Decision Tree Analysis, Commercial version of C4.5	Decision Tree	Windows, Unix
SIPINA	Provides an environment for supervised learning algorithms	C4.5, ID3	Windows, Linux
ORANGE	Provides data visualization and analysis for experts	Text mining and Bioinformatics add-ons	Windows, Linux
ALPHA MINER	Provides the best cost-and-performance ratio	Versatile data mining functions	Windows, Linux, Mac
WEKA	Provides machine learning algorithms for data mining tasks	Data pre-processing, classification, regression, clustering, association rules and visualization	Windows, Linux
Carrot	Provides ready-to-use components for fetching search results from various sources	Clustering	Windows, Linux
iData Analyzer	An Excel based data mining tool, Provides platform for visual learning environment	Pre processor, ESX, Heuristic Agent, Neural Network, Rule Maker and Report generator	Windows, Linux, Solaris

Table 2 : Commercial educational data mining tools

Name of Tool and developer	Function/Features	Techniques	Environment
Intelligent Miner (IBM)	Provides tight integration, Scalability of Mining	Association Mining, Classification, Regression, Predictive Modelling, Deviation detection, Clustering, Sequential Pattern Analysis	Windows, Solaris, Linux
MSSQL Server 2005	Provides DM functions in both relational database and data warehouse system.	Integrates the algorithms developed by third party retailers and application users	Windows, Linux
MineSet	Provides Robust Graphics tools	Association mining, Classification, advanced statics and visualization tools	Windows, Linux
Oracle Data Mining	Provides an embedded DWH infrastructure for multidimensional data analysis	Association mining, Classification, Prediction, Regression, Clustering, Sequence similarity search and analysis	Windows, Mac, Linux
SPSS Clementine	Provides an integrated data mining development environment for end users and developers	Association mining, Clustering, Classification, Prediction and visualization tools	Windows, Solaris, Linux
CART	Provides binary splitting and post pruning for Classification (Decision Tree) and for Prediction (Regression Trees)	Classification – Decision and Regression Tree	Windows, Linux

4. Literature survey:

Jasvinder Kumar [1] has presented Educational data mining as a new discipline in research community that applies various tools and techniques of data mining to survey data in the field of education. This paper combines all the modules of EDM like User and Stake holders of education, Tools, Techniques, Educational data, Task and Results. These modules are essential to facilitate the objectives of educational research.

B.Namratha, Niteesha Sharma [2] have discussed about educational data mining, its applications, benefits of educational data mining, challenges and barriers to successful application of educational data mining and learning analytics to improve teaching and learning. Data mining techniques in educational institutions help us to learn student performance, their behavior, designing course curriculum and to motivate students.

Ravi Tiwari, Awadhesh Kumar Sharma [3] have defined EDM as a discipline which deals with the application of the data mining tools and techniques to analyze the educational data. This discipline helps in the development of the models which facilitate effective decision making from the analysis of the historical educational data. It provides effective solution to various problems faced by the education sector. This paper describes various data mining tools and techniques used in educational data mining.

Dineshkumar B Vaghela, Priyanka Sharma [4] have discussed various data mining applications like Retail industries, Telecommunication industries, Financial data analysis, Biological data analysis, Intrusion detection etc. In this paper, they have presented new algorithm with Binary Search Tree which stores the global rules by consolidating the local rules generated at each site. This Global Rule Binary Search Tree (GRBST) can then be used in prediction of Students' admission to college.

Nikitaben Shelke, Shriniwas Gadge [5] has presented a survey of data mining approaches used for the performance analysis, prediction and evaluation. There are various tools and techniques for data mining in the development of educational and e-learning domain. Also, the summary of open source tools for data mining is presented.

Dr. P. Nithya, B. Umamaheswari, A. Umadevi [6] have described that Data mining is a vast area that includes different techniques and algorithms for finding patterns. The algorithms used in this paper have shown an improvement in strategies like course outline formation, understanding between teacher and student and high output and turn out ratio. This paper appreciates the existing algorithm researchers and instigates the new researchers.

Dr. Pranav Patil [7] has described a study of student's academic performance using data mining techniques. In this paper, we are discussing some research papers on the progress of students based on their performance. Most probably research work has been done on classification, clustering and association rules and for this WEKA data mining tool was mostly preferred for experiments.

5. CONCLUSION

There are various tools and applications of data mining in different fields. In this paper we have discussed applications of data mining in the field of education. Applications of data mining in the field of education are Analysis and visualization of data, Predicting Student Performance, Enrollment Management, Grouping Students, Predicting Student's Profiling, Planning and Scheduling, User Modeling, Organization of Syllabus, Detecting Cheating in Online Examination. We have also discussed, open source data mining tools and commercial data mining tools.

REFERENCES

[1] Jasvinder Kumar (2015), "A Comprehensive Study of Educational Data Mining", International Journal of Electronics & Computer Science Engineering Special Issue – TeLMISR, ISSN : 2348-2273.
[2] B.Namratha, Niteesha Sharma, "Educational Data Mining – Applications and Techniques", International Journal of Latest Trends in Engineering and Technology (IJLTET), ISSN: 2278-621X.
[3] Ravi Tiwari, Awadhesh Kumar Sharma (2015), "A Detailed Analysis of Educational Data Mining", International Journal for Research in Emerging Science and Technology, Volume-2, Issue-7.
[4] Dineshkumar B Vaghela, Priyanka Sharma (2015), "Students' Admission Prediction using GRBST with Distributed Data Mining", Communications on Applied Electronics (CAE), Volume 2.
[5] Nikitaben Shelke, Shrinivas Gadge (2015), "A Survey of Data Mining Approaches in Performance Analysis and Evaluation", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, ISSN: 2277 128X.
[6] Dr. P. Nithya, B. Umamaheswari, A. Umadevi (2016), "A Survey on Educational Data Mining in Field of Education", Journal of Computer Science and Software

Development, Volume 1, Issue 1.
[7] Dr. Pranav Patil, "A Stusy of Student's Academic Performance Using Data Mining Techniques", International Journal Of Research In Computer Applications And Robotics, ISSN 2320-7345.